

Colegio Universitario de Estudios Financieros
Máster en Data Science



TRABAJO DE FIN DE MÁSTER

**ANÁLISIS DE LA MOVILIDAD PEATONAL EN LA CIUDAD DE MADRID:
ANÁLISIS EMPÍRICO Y LEYES DE POTENCIA**

AUTOR: Alejandro García Girón

DIRECTORES:

Prof. María Victoria López López

Prof. José María Sarabia Alegría

SEPTIEMBRE 2021

AGRADECIMIENTOS

Quiero agradecer a mis directores del Trabajo Fin de Máster, la profesora López y el profesor Sarabia la dedicación e implicación que han tenido para que este trabajo haya sido posible. A mis Padres, a mi abuela Carmen y a mis tíos Alfonso y Cristina, Carmen y Fernando por todo su apoyo, confianza y cariño durante estos años de trabajo.

Quiero agradecer a mis compañeros del Máster en *Data Science* por su cooperación y compañerismo que han tenido ayudando y aconsejándome en cada una de las labores académicas que he llevado a cabo, en especial a Gabriel Blanco, Octavio del Sueldo, Francisco del Val, Laura Martínez, Matias Corredoira, Beatriz Quevedo, Pablo Revuelta, Luis Vaciero y Antonio Romero.

A mi estimado amigo Adrián Rodríguez Caballero por darme su visión optimista de todo problema y su calidad de la amistad.

A la profesora M^a Jesús Mures Quintana por hacerme entender la estadística de una manera sencilla y aplicada, lo que ha servido para que mi especialización académica haya tomado ese rumbo.

A José Rubén Gallardo Gómez, por haber confiado en mí en mi primera experiencia laboral, lo cual estaré agradecido siempre.

En último lugar, quiero agradecerme a mí mismo el hecho de confiar en mí.

RESUMEN

Esta memoria de Fin de Máster consiste en una investigación para demostrar el cumplimiento de la llamada *Power Law*, haciendo uso de los datos de movilidad de la ciudad de Madrid para el año 2020. Se ha elegido los datos correspondientes al distrito Centro, tomando como referencia de la investigación la movilidad peatonal recogida por los sensores distribuidos por la capital de España.

Para abordar este trabajo se ha distribuido el contenido en seis capítulos. En primer lugar, se introduce la problemática y la casuística sobre la que trabajaremos junto con los objetivos que perseguimos cumplir. En segundo lugar, se ha trabajado sobre memorias y publicaciones referentes sobre esta temática de distintas épocas, comparando las distintas metodologías de ajuste que emplean los autores para alcanzar el objetivo final de validación de la *Power Law*. En tercer lugar, se han aplicado dichas metodologías al conjunto de datos, haciendo uso de los consecuentes contrastes y validaciones. Por último, se ha incluido en un anexo las notas relevantes y conclusiones extraídas durante la investigación.

En la memoria se persigue un doble objetivo: por un lado, el procesado, análisis y limpieza de los datos y por otro el de probar que la movilidad en la ciudad de Madrid sigue una ley de potencias, como designaremos más frecuentemente, *Power Law*.

Palabras Clave: Sensores, movilidad, periodo de confinamiento, modelo tamaño-rango, validación, regresión tamaño-rango $\frac{1}{2}$.

ABSTRACT

This Master's End report consists of an investigation to demonstrate compliance with the Power Law. with the mobility data of the city of Madrid in the year 2020, particularly of the Centro district, taking as a reference in the research the pedestrian mobility collected by the sensors distributed by the capital of Spain.

To address this work, the content has been distributed into six chapters. In a first step the problem and the casuistry on which we will work together with the objectives that we seek to achieve are introduced. In a second step, we have worked on memoirs and referential publications on this subject of different periods comparing the different adjustment methodologies used by the authors to achieve the final objective. In a third step, these methodologies have been applied to the data set together with the respective validations and contrasts. Finally, an appendix with the relevant comments and conclusions drawn during the investigation is included.

A double objective is pursued, on the one hand, the processing, analysis and data cleansing of the data and in other hand, to prove that mobility in the city of Madrid follows a power curve, as we will more frequently designate, Power Law.

Key Words: Sensors, mobility, confinement period, size-range model, validation, size-range regression $\frac{1}{2}$.

ÍNDICE

1.	INTRODUCCIÓN	1
2.	OBJETIVOS	2
3.	REVISIÓN DE LA LITERATURA	3
4.	METODOLOGÍA.....	6
4.1.	FUENTES DE INFORMACIÓN	6
4.1.1	Estructura de los datos.....	6
4.1.2	Transformación de los datos.....	7
4.2.	MODELO POWER LAW	8
4.2.1	Metodología tamaño-rango.....	8
4.2.2	Casuística en la naturaleza.....	11
5.	ANÁLISIS EMPÍRICO	12
5.1	ANÁLISIS DESCRIPTIVO DE LOS DATOS.....	12
5.2	AJUSTE DE LOS DATOS	15
5.2.1	Ajuste año 2020	15
5.1.2	Ajuste del período de confinamiento	19
5.1.3	Ajuste del período estival.....	23
5.	CONCLUSIONES Y TRABAJO FUTURO	27
6.	REFERENCIAS.....	28
	ANEXO	30

ÍNDICE DE TABLAS

Tabla 4.1 Lista de variables.....	7
Tabla 5.1.1 Estadísticos descriptivos.....	13
Tabla 5.1.2 Deciles del volumen de peatones 2020.....	13
Tabla 5.1.3 Deciles del volumen de peatones , período de confinamiento.....	14
Tabla 5.1.4. Deciles del volumen de peatones , período estival.....	14

ÍNDICE DE FIGURAS

Figura 4.2 Distribución Power Low.....	10
Figura 5.1.1. Volumen de peatones en el distrito Centro de la ciudad de Madrid.....	12
Figura 5.2.1 Representación del logaritmo del volumen de peatones y del rango para el período 2020.....	15
Figura 5.2.2. Salida del análisis de regresión tamaño-rango para la variable logaritmo del rango (x) frente al logaritmo del tamaño, número de peatones en el año 2020.....	16
Figura 5.2.3. Representación de los valores de R2 para el período 2020.....	16
Figura 5.2.4. Ajuste del subconjunto por el método de MCO.....	17
Figura 5.2.5. Regresión óptima tamaño-rango para el año 2020.....	18
Figura 5.2.6 Ajuste del subconjunto por el método Gabaix e Ibragimov para el periodo 2020.....	18
Figura 5.2.7 Representación del logaritmo del volumen de peatones y del rango para el el periodo de confinamiento.....	19
Figura 5.2.8. Salida del análisis de regresión tamaño-rango para la variable logaritmo del rango (x) frente al logaritmo del tamaño, número de peatones para el periodo de confinamiento.....	20
Figura 5.2.9. Representación de los valores de R2 para el período de confinamiento.....	20
Figura 5.2.10. Ajuste del subconjunto del confinamiento por el método de MCO.....	21
Figura 5.2.11. Regresión óptima tamaño-rango para el periodo de confinamiento.....	21
Figura 5.2.12 Ajuste del subconjunto por el método Gabaix e Ibragimov para el periodo de confinamiento.....	22
Figura 5.2.13 Representación del logaritmo del volumen de peatones y del rango para el el período estival.....	23
Figura 5.2.14. Salida del análisis de regresión tamaño-rango para la variable logaritmo del rango (x) frente al logaritmo del tamaño, número de peatones para el periodo estival.....	23
Figura 5.2.15. Representación de los valores de R2 para el período estival.....	24
Figura 5.2.16. Ajuste del subconjunto estival por el método de MCO.....	24
Figura 5.2.17. Regresión óptima tamaño-rango para el periodo estival.....	25
Figura 5.2.18 Ajuste del subconjunto por el método Gabaix e Ibragimov para el periodo estival.....	25

1. INTRODUCCIÓN

La movilidad peatonal en las ciudades es un fenómeno sobre el cual se han realizado diversos estudios, por ejemplo, el de Xin et al (2014). Estos autores analizan la movilidad mixta, peatones y tráfico. Generalmente muchos de los estudios en este campo se centran en la movilidad de vehículos, y tienen como objetivos el optimizar los movimientos en vehículos, minimizar atascos, además de, en multitud de ocasiones, tratar de reducir las emisiones de carbono, riesgos de accidente y realizar previsiones de movilidad. El presente trabajo presenta sin duda un carácter de trabajo de investigación, donde se produce una unión entre la búsqueda de los datos, y el contraste y validación de la ley de potencias, probando que la movilidad acumulada por horas sigue dicha ley de potencia.

La ciudad de Madrid, de acuerdo con su padrón municipal del año 2020, cuenta con una población, según la información del Instituto de Estadística (2021), de 3.334.730 habitantes. Dichos habitantes se reparten en 21 distritos, entre los cuales se ha seleccionado el distrito Centro, que es uno de los más transitados y cuenta con once cámaras de medición de actividad peatonal. Para medir el flujo peatonal el Ayuntamiento de Madrid ha dispuesto sensores en las calles de la ciudad que realizan conteo continuo del paso de peatones. Concretamente, en el distrito Centro de Madrid los sensores se encuentran en las calles de San Bernardo, Fuencarral, Hortaleza, San Jerónimo, Bailén, Atocha, Mayor, Huertas, Recoletos, Gran Vía y Alcalá.

El objetivo del presente trabajo es probar si existe relación entre la movilidad peatonal en el distrito Centro de la ciudad de Madrid y la distribución de probabilidad del tipo *Power Law*, también conocida como ley de potencias. Para ello, se ha extraído la información del portal de datos abiertos del Ayuntamiento de Madrid relativa a movilidad urbana para el año 2020.

La estructura de la memoria es la siguiente. En primer lugar, hablaremos de los objetivos que se persiguen. A continuación, se revisa la literatura más destacada sobre esta temática y después explicaremos la metodología probabilística e inferencial utilizada, así como el tratamiento de los datos extraídos. En tercer lugar, se realizará un análisis empírico de los datos, tanto gráfico como analítico, para finalmente comentar las principales conclusiones.

2. OBJETIVOS

Durante los últimos años, y más concretamente en el último lustro, ha habido multitud de estudios basados en la búsqueda de relaciones entre fenómenos reales y la distribución *Power law* o ley de potencias. La ley de potencias, explica un buen número de situaciones y confirma relaciones propias de la naturaleza tal como argumentaba Klaus et al. (2011) en relación con las avalanchas neuronales. En este trabajo, el aspecto que se investiga se refiere al de datos relativos a la movilidad peatonal, la cual ya fue estudiada por Xu et al. (2019), aunque esta última tenía un enfoque más dirigido al tráfico mixto entre peatones y vehículos.

El objetivo principal de esta investigación es demostrar que la movilidad peatonal del distrito Centro de la ciudad de Madrid sigue una distribución de tipo *Power Law*. Para poder materializarlo, necesitaremos obtener los datos y hacer un preprocesado de los mismos. Este preprocesado incluye: limpieza, transformación y determinar si son válidos para conseguir el propósito de este trabajo. Una vez que los datos están en disposición de modelizarlos se buscará la relación lineal de la movilidad, representando los datos en acumulados por horas, ordenados y transformados en logaritmos en el eje de abscisas y en el eje de ordenadas el logaritmo del rango que ocupan los mismos, de la misma forma que realizaron Yamamoto y Miyazima (2004). Estos autores trabajaron con los ingresos muestrales de la población de asalariados japoneses, de modo que el diagrama finalmente obtenido será una recta con pendiente negativa. En cuanto a la validación del modelo, hay autores que prefieren realizar el correspondiente contraste con una metodología basada en estimación por máxima verosimilitud, complementada con ratios de probabilidad como argumenta Clauset et al. (2009). Otra segunda alternativa, es utilizar mínimos cuadrados ordinarios sobre las variables rango y tamaño (en nuestro caso movilidad). La diferencia entre las dos metodologías es mínima, de modo que Clauset et al. (2009) también la emplea. En el presente trabajo se hará un doble contraste. Haremos uso de la metodología de Clauset et al. (2009), así como de la propuesta de Gabaix e Ibragimov. (2011), que es una variación de la metodología propuesta por Clauset et al. (2009).

3. REVISIÓN DE LA LITERATURA

En este capítulo se realiza una revisión de la literatura y se muestran casos en los que se cumple de forma natural la Ley de potencias citada anteriormente. Algunos autores realizan precisiones para demostrar el ajuste de esta ley como se describe a continuación.

En la investigación titulada "*The power-law distribution of cumulative coal production*" de Balthrop y Quan (2019), los autores realizan un estudio sobre la producción de carbón acumulada y su ajuste a la Power Law en los Estados Unidos. El estudio demuestra que los datos se ajustan a esta distribución de probabilidad *Power Law*. Los autores toman como coordenadas el nivel de producción de la mina y el rango distributivo de la misma. Mediante contrastes de bondad del ajuste, se concluye que esta distribución se aproxima mejor a los datos que otras distribuciones alternativas, incluso mejorando el de la ley Log-Normal. En un principio se conocía que esta distribución explicaba ciertos fenómenos sociales y económicos como el tamaño de las ciudades o el índice de popularidad. Sin embargo, cada vez son más los fenómenos que son explicados por este tipo de distribución.

En el estudio de Clauset et al. (2009), titulado "*Power-law distributions in empirical data*", los autores estudian una serie de veinticuatro fenómenos naturales recogidos en varios conjuntos de datos donde, a priori, se conjetura que pueden seguir una *Power Law*. En el trabajo se explora como se distribuyen todos estos conjuntos de datos. El autor replantea su estudio con un enfoque basado en la dificultad de identificar estas distribuciones, debido a que se trata de distribuciones con colas pesadas y de identificar el rango sobre el que se comporta la ley. Los autores también sostienen que el método de mínimos cuadrados no es totalmente preciso para identificar estas leyes, y en cambio apuestan por un enfoque basado en métodos de máxima verosimilitud y contrastes de bondad Kolmogorov- Smirnov basados en metodologías Bootstrap. Se concluye que la linealidad de puntos entre dos ejes logarítmicos es una condición necesaria pero no suficiente para demostrar que ese subconjunto de datos sigue el comportamiento de una *Power Law*.

Los autores Xin et al. (2014), trabajaron la movilidad mixta en la publicación, "*Power-law in pedestrian crossing flow under the interference of vehicles at an unsignalized midblock crosswalk*", la cual quizás sea la que guarde más similitud con la presente investigación, puesto que la temática es el análisis del tráfico. Los autores buscan un modelo para explicar la movilidad mixta, tanto vehicular como peatonal. Entre los diversos objetivos que se plantean, consideran el encontrar patrones de comportamiento en función de las características y calidad de los movimientos y en consecuencia maniobras viales de posible riesgo. Todo esto pretende evitar problemas de seguridad vial, que afectan continuamente al tráfico en las ciudades. Los investigadores reconocieron varias fases en el flujo del tráfico: la fase libre, la fase libre de peatones, la fase sin vehículos y la fase de atascos.

La mayoría de los peatones cruzan las vías en pequeños grupos y una minoría lo hace en grandes grupos. El tamaño de estos grupos que cruzan las calles sigue un comportamiento común tipo *Power Law*, lo que sugiere una cuantificación del fenómeno natural y estocástica. También llegaron a concluir que en las ocasiones donde el tráfico es abundante, solo siguen distribución *Power Law* aquellos conglomerados donde el sistema los califica de grupos heterogéneos, dejando de lado aquellas situaciones donde los grupos de peatones son más homogéneos.

Los autores Xu et al. (2019), en la publicación, "*Understanding the diversity on power-law-like degree distribution in social networks*", abordan un estudio de las relaciones interpersonales que tienen los individuos en las redes, basándose en la premisa de que existen algunos individuos con muchas relaciones con otras personas y muchos individuos donde las relaciones son sustancialmente inferiores. Los autores investigan el orden de estas interacciones bidireccionales y proponen un modelo. Los resultados son que el modelo reproduce la distribución de dichas relaciones y son similares a una distribución *Power-Law*. Además, visualizan que el factor tiempo de uso y la jerarquía social de los individuos son la razón que justifica esta distribución de las relaciones y su diversidad.

Los autores Klaus et al (2011) abordan una visión aplicada al campo de la salud en su publicación "*Statistical analyses support power law distributions found in neuronal avalanches*". Estos autores parten de que la distribución del tamaño de las avalanchas neuronales en las redes corticales sigue una ley de potencia. La identificación de la escala de las mismas es una cuestión sustancialmente más compleja, sobre todo cuando se trabaja con los datos disponibles.

Yamamoto y Miyazima (2004) en su investigación "*Power-law behavior in social and economical phenomena*", se centran en la demostración de esta ley en el contexto de los escenarios económicos. Los autores sugieren un modelo según el cual se distribuye el ingreso económico de las empresas basándose en la fractalidad del mercado, es decir, en que los patrones relacionales de las actividades son similares en todos los escalones del sistema financiero. Este es el principal razonamiento que da lugar al uso de la distribución *Power Law*. El estudio recoge datos de los asalariados de Japón y los representa en un plano doblemente logarítmico. Se propone un diagrama con un eje donde son representados el importe de los ingresos y en el otro eje se clasifican estos mismos sin ningún parámetro de ajuste. Posteriormente diseñan flujos de variación de riqueza y generan clasificaciones basadas en arboles de decisión.

Lorenzo et al. (1990) en el artículo, "*Escalamiento de magnitud en la satisfacción laboral: I. Validez de la ley de potencias*", estudian la ley de potencias en el ámbito laboral. En el artículo los investigadores quieren determinar si la satisfacción laboral medida sobre los empleados de Correos y Telégrafos sigue una *Power Law*. Previamente, preparan los datos con un escalamiento directo de las opiniones recogidas mediante la técnica de Stevens. La recogida de la información se basó en un cuestionario donde los muestreados respondían a un preguntas multi-respuesta, que posteriormente se escalaban y ponderaban. Las respuestas son proyectadas en graficas con el logaritmo de

la ponderación y las distintas categorías. Los autores no emplean una metodología doblemente logarítmica en las proyecciones, sino que consideran solamente un eje numérico y el otro categórico. El análisis concluye interpretando la distribución de las respuestas con correlaciones, más que un análisis contrastado de la distribución estadística de los mismos.

Un enfoque en el ámbito económico fue proporcionado por Gabaix (2016), en su artículo *"Power Law in Economics: An Introduction"*. En dicho trabajo hace una introducción sobre esta ley y sobre sus aplicaciones en el campo de la economía. Afirma que puede parecer trivial su identificación, ya que muchos fenómenos de la economía (como por ejemplo la curva de demanda) tienen pendiente negativa, al igual que ocurre con la ley de potencias. Gabaix estudia esta ley con una muestra de ciudades de EE. UU., donde proyecta los puntos del logaritmo de los habitantes y logaritmo del rango y demuestra en este campo la ley de potencia. El autor también demuestra que los rendimientos de las acciones del mercado de valores siguen una ley de potencias, ya que muestran unas colas pesadas y al aplicarlas las transformaciones logarítmicas se demuestra que siguen este comportamiento de *Power Law*. Entre las conclusiones se obtiene que en los macrodatos que genera la economía, estas relaciones están muy presentes. Por otro lado, el hecho de que los datos presenten colas pesadas puede dar pistas al investigador sobre la presencia de este fenómeno.

Los autores Gabaix e Ibragimov (2011) en su publicación *"Rank- $\frac{1}{2}$: a simple way to improve the OLS estimation of tail exponents"*, estudian una modificación de la regresión habitual basada en la diferencia de medio punto en el rango, de tal forma que especifican el modelo,

$$\log \left(\text{Rango} - \frac{1}{2} \right) = a - b \log (\text{Tamaño}).$$

La justificación de esta nueva regresión es que el procedimiento está fuertemente sesgado en muestras pequeñas. Esta metodología está considerada como un remedio práctico simple para mitigar el sesgo. Los autores concluyen sobre el número de observaciones óptimas a la hora de llevar a cabo estimaciones en las colas de la distribución.

4. METODOLOGÍA

Este capítulo se divide en dos secciones. En primer lugar, se explican las fuentes de datos y el tratamiento necesario para su análisis posterior. En segundo lugar, se tratan los aspectos estadísticos. Posteriormente en el capítulo 5 aplicaremos la metodología *Power Law* a los datos de movilidad.

4.1. FUENTES DE INFORMACIÓN

El Ayuntamiento de Madrid en su calidad de gobierno abierto, pone a disposición de los ciudadanos un portal de datos abiertos con un catálogo muy amplio. A fecha de la escritura de esta memoria se pudo acceder a un total de 503 conjuntos de datos en diversos formatos (csv, xlsx, geo, json, html...). De este catálogo se ha descargado el fichero csv donde se incluyen los datos de movilidad para el período 2020.

4.1.1 Estructura de los datos

Los datos vienen acompañados de una guía donde se muestra la estructura de los mismos, características y formato de las mediciones. El conjunto de datos se denomina “*Aforos de peatones y bicicletas*” y se encuentra disponible en el portal de datos abiertos del Ayuntamiento de Madrid. Los datos son recogidos por 40 sensores automáticos de movilidad peatonal y ciclista. Además, existen 120 de tráfico vehicular. La recogida se realiza mediante videocámaras, sensores convencionales y sistemas de identificación de móviles, de modo que se puede conocer la dirección y volumen del tráfico peatonal y de bicicletas. Las balizas están repartidas a lo largo del municipio. Según Ticpymes (2019), dichas balizas tienen el fin de capturar, agregar, archivar y supervisar diversas fuentes de datos, ayudando a los técnicos municipales a obtener una vista agregada de los mismos para su posterior análisis.

La recogida se realiza en intervalos de una hora y se almacena con una periodicidad de 24 registros diarios por sensor. La unidad responsable de la captación y archivo de los datos es la Dirección General de Planificación e Infraestructuras de Movilidad del Ayuntamiento de Madrid.

Las variables que definen cada una de las observaciones se recogen a continuación.

Nombre del campo	Tipo	Descripción
FECHA	string	Fecha de recogida de datos
HORA	string	Hora de recogida de datos
IDENTIFICADOR	string	Identificador del aforo. Indicador del número de estación Permanente de aforo, tanto de Peatón como de Bicicleta
BICICLETAS/PEATONES	integer	Número de peatones o bicicletas contadas en un cuarto de hora
NÚMERO_DISTRITO	integer	Número del distrito donde se encuentra la estación de aforo
DISTRITO	string	Denominación del Distrito donde se encuentra la estación de aforo
NOMBRE_VIAL	string	Calle donde se encuentra la estación de aforo
NÚMERO	integer	Número de la calle donde se encuentra la estación de aforo
CÓDIGO_POSTAL	integer	Código postal donde se encuentra la estación de aforo
OBSERVACIONES_DIRECCION	string	Información adicional a la dirección
LATITUD	float	Latitud en coordenadas geográficas
LONGITUD	float	Longitud en coordenadas geográficas

Tabla 4.1. Lista de variables, Fuente: Portal de datos Ayuntamiento de Madrid

Los datos se presentan en tres formatos; formato *String*, son las cadenas de texto; *integer*, es el formato que presentan los datos enteros (sin decimales, como el número de peatones que circulan en un segmento temporal) y formato *float*, que en este caso son las coordenadas geográficas, latitud y longitud, donde se sitúan los sensores de medición.

4.1.2 Transformación de los datos

Los datos, tal como se comprobará en la sección 4.1, son claramente la materia prima de la investigación, los cuales se han extraído de la web del Ayuntamiento de Madrid. Se ha descargado un fichero con formato csv donde se recogen 362.213 observaciones acompañados de 12 columnas. Dichas columnas recogen la información correspondiente a cada sensor de movimiento: número de peatones, código postal, latitud, longitud, nombre de la vía, distrito, observaciones particulares, hora, fecha y

número de distrito. Se ha segmentado el año en dos períodos para estudiar la particularidad y el impacto de la pandemia Covid-19, ya que las restricciones de movilidad peatonal fueron notorias, así como la movilidad en los meses de verano.

Lo tramos analizados son, el período 2020 en general, creando una acumulación de los datos de todo el año para su contraste con la distribución *Power Law*. Por otro lado, contrastamos el período de confinamiento de Covid-19 cuya medición es del 15 de marzo al 15 de mayo del año 2020. Finalmente, la época estival, es decir, los meses de julio y agosto del mismo año. En esta última se reduce sustancialmente el flujo peatonal en Madrid, concretamente en agosto.

Los datos se cargan en el software R Studio donde se limpian, modelan y grafican. El primer paso ha sido la limpieza de datos, dónde se visualizó el número de observaciones nulas, que ha sido del 53%, un número bastante alarmante, después de agrupar los datos por distritos. Entonces, determinamos que el distrito que vamos a trabajar, Centro, tiene un 0% de valores nulos. Una vez limpios y filtrados por distrito, acumulamos la variable número de peatones por horas y los representamos gráficamente. En el eje de ordenadas el logaritmo del rango del volumen peatones por hora, en orden ascendente, y en el eje de abscisas el logaritmo del volumen o tamaño de peatones.

La validez la estudiaremos gráfica y analíticamente, para confirmar que los datos siguen una *Power Law*. Analizaremos el contraste global del modelo con el estadístico F y la capacidad explicativa del mismo con el coeficiente R^2 y R^2 ajustado. En el plano gráfico buscaremos que la relación lineal esté en el último tramo de la pendiente que es donde se cumple esta relación, es decir, en la parte alta de la distribución.

4.2. MODELO POWER LAW

4.1.3 4.2.1 Metodología tamaño-rango

Según Yaneer Bar-Yam (2011), una ley de potencia es una distribución que se basa en una relación en la que un cambio relativo en una cantidad genera un cambio que es proporcional a la cantidad que variaba inicialmente. Todo ello de manera independiente al tamaño inicial de las cantidades estudiadas.

Para llevar a cabo la investigación seguiremos la metodología de ajuste-validación de la Power-Law basado en Clauset et al. (2009) y en Gabaix y Ibragimov. (2011) la cual podemos sintetizar en los siguientes pasos:

- Análisis gráfico previo del total de los datos Tamaño-Rango.
- Selección óptima del número de datos a partir del gráfico (k , R^2), donde k es número de observaciones prescindiendo de las k -mayores ($k=1, 2, \dots$) y R^2 es el coeficiente de determinación de la regresión tamaño-rango. Se elige el valor de k con mayor R^2 (máximo local global).

- Ajuste mediante MCO de los datos tamaño-rango (en logaritmos) basado en el número óptimo de puntos obtenidos en el paso anterior.
- Robustez de los resultados obtenidos en el paso 4, mediante la regresión tamaño-rango-1/2, basado en Gabaix e Ibragimov (2011).
- Validación del modelo obtenido mediante el p-valor del coeficiente beta y el p-valor de la F global del modelo.

El término robustez se refiere a que, si ambas regresiones coinciden el ajuste es robusto y los resultados obtenidos son los definitivos. Sin embargo, si hay diferencias importantes en la pendiente, nos quedaríamos con la segunda de las regresiones, puesto que es insesgada en pequeñas muestras.

El método de Gabaix y Ibragimov. (2011) es una variación del método propuesto por Clauset et al. (2009) y ampliamente usado en la literatura. Estos autores han observado que el ajuste tamaño-rango es sesgado frente a pequeñas muestras. Para corregir este efecto no deseado, se introduce la regresión tamaño-rango-1/2.

El procedimiento que seguiremos en el presente estudio es el que presentamos a continuación, donde partimos de una variable aleatoria, número de peatones, que está definida mediante la función de distribución de tipo continuo,

$$F(x) = \Pr(X \leq x) = 1 - \left(\frac{x}{x_0}\right)^{-\alpha}, x \geq x_0 > 0, \quad (1)$$

donde $\alpha > 0$ y $F(x) = 0$ si $x \leq x_0$.

Escribimos:

$$1 - F(x) = \Pr(X \leq x) = \left(\frac{x}{x_0}\right)^{-\alpha} \quad (2)$$

y tomando logaritmos en ambos lados se obtiene:

$$\log(1 - F(x)) = -\alpha \log x + \alpha \log x_0, \quad (3)$$

donde $x \geq x_0$. Si disponemos de un conjunto de datos, la ecuación (3) es equivalente a considerar el modelo:

$$\log(R(x_i)) = \beta_0 + \beta_1 \log(x_i) \quad (4)$$

donde $R(x_i)$ es el rango de datos i -ésimo, $x_i \geq x_0, i = 1, \dots, n$, siendo x_0 el umbral.

El valor de β_1 debe ser negativo para que se cumpla la ley. Los casos relevantes son:

- Sí $\beta_1 = -1$ tenemos la ley de Zipf
- Sí $\beta_1 \neq -1$ tenemos ley de potencias

Para el ajuste del modelo, tendremos el plano gráfico donde dibujamos los puntos los puntos $(\log(x_i), \log(R_i))$, $i = 1, 2, \dots, n$. Se debe obtener una recta con pendiente negativa.

En la práctica podemos ajustar el modelo (4) mediante regresión lineal (MCO):

$$\log(R(x_i)) = \beta_0 + \beta_1 \log x_i + u_i \quad (5)$$

donde u_i es el término de error de la regresión. Una vez estimada la regresión, para que se satisfaga la ley de potencias ha de cumplirse:

- El modelo ha de ser lineal. Esto puede ser contrastado por el estadístico F de Snedecor
- Al tratarse de un modelo de regresión simple, lo anterior es equivalente a que el coeficiente β_1 sea significativo, para lo que planteamos el contraste $H_0: \beta_1 = 0$ frente a $H_1: \beta_1 \neq 0$ usando el estadístico t.

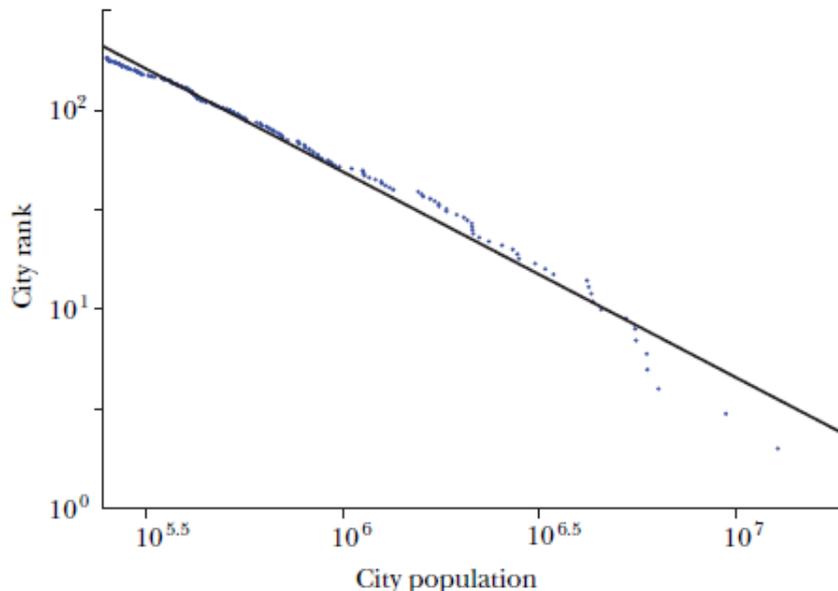


Figura 4.2. Distribución Power-Law. Fuente: Gabaix e Ibagimov (2011: 187)

En el gráfico 4.2 se muestra el ajuste de los datos en escala doblemente logarítmica que propusieron Gabaix e Ibragimov. (2011).

4.1.4 4.2.2 Casuística en la naturaleza

En la naturaleza hay muchos fenómenos cuyas observaciones siguen patrones del tipo Power Law. Clauset et al (2009), citan una serie de ejemplos como: la cantidad de llamadas recibidas por los clientes de una compañía telefónica en Estados Unidos a lo largo de un día, el número de muertes por batalla en las guerras del siglo XIX, el número de bytes que arrojan como resultado las solicitudes a una web http en un período de 24 horas, entre otros fenómenos. En otros campos científicos de estudio como la física, la fuerza y el potencial en el movimiento armónico simple siguen este tipo de patrones. También se cumplen en la biología, la astronomía, las matemáticas o las finanzas.

5. ANÁLISIS EMPÍRICO

En este capítulo se analizan los datos recogidos y se les aplicará la metodología que expusimos en el punto 4 del presente trabajo. En primer lugar, valoraremos un análisis descriptivo de los datos con los principales descriptivos de dispersión, posición y forma. A continuación, analizamos todas las observaciones que pertenecen al año 2020, así como dos segmentos temporales, el período de confinamiento que sufrió España en el año 2020 y los meses estivales (julio y agosto).

5.1 ANÁLISIS DESCRIPTIVO DE LOS DATOS

En este apartado vamos a desarrollar los principales descriptivos de dispersión posición y forma, además de gráficos de movilidad.

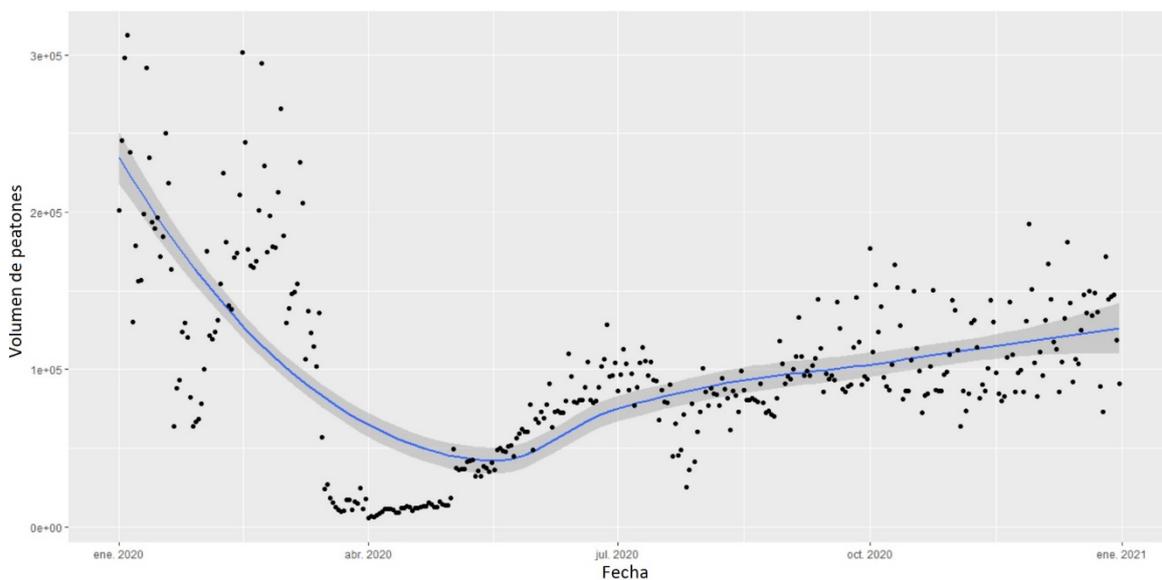


Figura 5.1.1. Volumen de peatones por mes en el distrito Centro de la ciudad de Madrid. Fuente: Elaboración propia

Como se aprecia en el gráfico 5.1.1 el descenso de movilidad se vio afectado en marzo de 2020 con la imposición del confinamiento debido a la pandemia causada por el Covid-19. En la gráfica se observan los dos subperiodos que se tratan, el descenso de movilidad y el crecimiento paulatino de la misma que no alcanza de nuevo los volúmenes de movilidad que había antes del confinamiento a comienzo del año 2020.

Los estadísticos descriptivos que se van a analizar son los correspondientes a las variables de volumen de peatones por hora y día de los períodos estudiados. Se incluirán gráficos de la evolución de la movilidad.

	Min	Mediana	Media	Varianza	Desviación	Max
Año 2020	0	80	314.6278	626212.65	791.336	16673
Verano 2020	0	77	262.4011	298380.31	546.242	6735
Período confinamiento 2020	0	20	62.2682	35221.91	187.675	9749

Tabla 5.1.1. Estadísticos descriptivos. Fuente: Elaboración propia

Para el año 2020, como se ve en la tabla 5.1.1, el número máximo de peatones diarios que circularon por uno de los sensores de las calles del distrito Centro en una hora de un día fue de 16.673 peatones y un mínimo de 0 peatones. La media, tomada a partir de los registros por hora es de 314 peatones, en las horas de la madrugada el número de peatones suele presentar frecuencia 0. La mediana, que es el punto donde los extremos de la distribución se presentan equidistantes es de 80 peatones, por lo que la frecuencia de observaciones donde el número de peatones es igual a 0 es alto. La desviación típica es de 791, sustancialmente alta respecto a la media lo que nos viene a decir que no es representativa. Lo comprobamos con el coeficiente de variación que es 2,5 que por ende al superar el indicador de 0,3, es decir, el 30%, concluimos que los datos son heterogéneos. Respecto a la posición de los datos tenemos la siguiente distribución tomada por deciles.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0	6	18	32	47	80	127	199	334	763	16673

Tabla 5.1.2. Deciles del volumen de peatones 2020. Fuente: Elaboración propia

En el gráfico anterior, Tabla 5.1.2, se ve como la distribución de los datos se concentra mayormente en la parte alta, teniendo en cuenta que el distrito centro es uno de los que más tráfico peatonal tiene de la capital de España. Las medidas de forma presentan un coeficiente de asimetría de 7,46 y de curtosis de 90,15 lo que presenta una asimetría positiva y una forma leptocúrtica.

En el período de confinamiento, el cual hemos acotado en un segmento bimestral comprendido entre el 15 de marzo de 2020 y el 15 de mayo del mismo año, presentamos un claro descenso en el número medio de peatones teniendo una media de 62 peatones, una mediana de 20 y un máximo de 9749 peatones. La media tampoco es representativa ya que el coeficiente de variación es de 3,01 por lo que los datos no son homogéneos. La distribución de los datos es la siguiente.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0	1	5	10	17	20	33	40	64	121	9749

Tabla 5.1.3. Deciles del volumen de peatones período confinamiento. Fuente: Elaboración propia

En la tabla 5.1.3 se aprecia que la mayoría del volumen se concentra en la parte alta de la distribución, concretamente a partir del decil 9. Los coeficientes de asimetría y curtosis son de 16,8 y de 570,3 lo que nos indica una asimetría positiva y una forma leptocúrtica.

En el período estival, julio y agosto, la media de peatones fue de 262 con una mediana de 67. El número máximo fue de 6735 y el coeficiente de variación de 2,08 lo que confirma que los tres conjuntos de datos son bastante heterogéneos. La distribución por deciles se presenta a continuación.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0	6	18	35	45	77	117	179	290	679	6735

Tabla 5.1.4. Deciles del volumen de peatones en el período estival. Fuente: Elaboración propia

Como se observa en el gráfico anterior, tabla 5.1.4, se aprecia que la distribución coje forma al final de esta donde acumula la mayor cantidad de peatones. Los tres subsegmentos temporales siguen los mismos patrones de distribución. El coeficiente de asimetría es de 4,43 por lo que es asimétrica positiva y la forma es leptocúrtica con un coeficiente de 28,21, algo menos apuntada que las anteriores.

5.2 AJUSTE DE LOS DATOS

En esta fase se va a analizar el periodo del año 2020 junto con los dos subperiodos dentro de este año que definimos anteriormente, con la metodología clásica de ajuste y con la metodología de Gabaix e Ibragimov (2011).

4.1.5 5.2.1 Ajuste año 2020

En el gráfico que se presenta a continuación se representa el volumen de patones acumulado por horas en una doble escala logarítmica que se detalla a continuación.

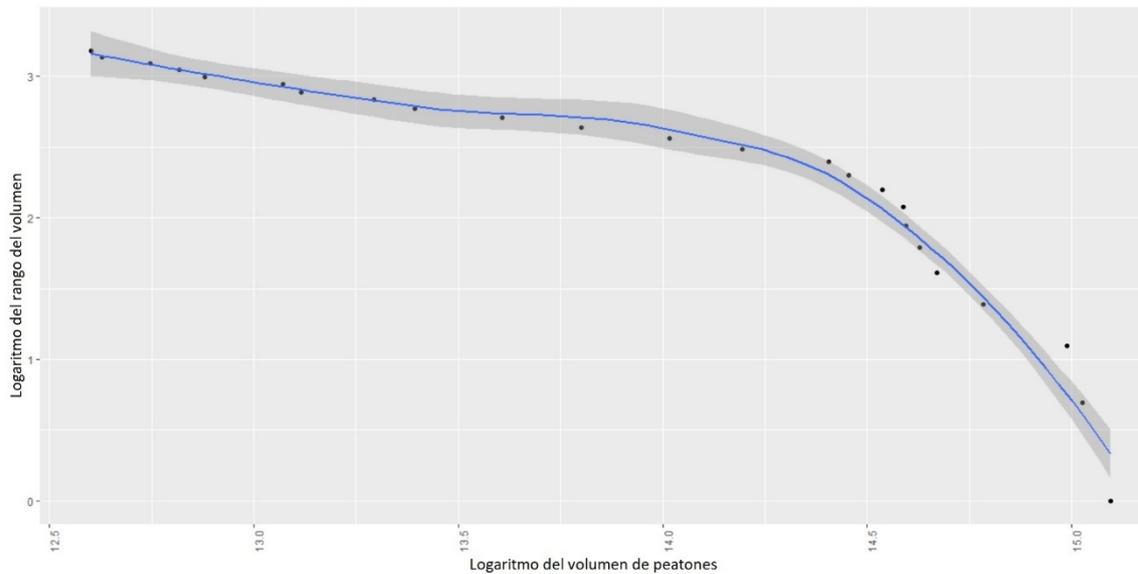


Figura 5.2.1 Representación del logaritmo del volumen de peatones y el del rango para el periodo 2020.
Fuente: Elaboración propia

En el gráfico anterior, figura 5.2.1, en el eje de abscisas se puede observar el logaritmo del número de peatones que circularon por el distrito Centro en el año 2020. En el eje de ordenadas se representa el Rango, $R(x_i)$, de la escala ordinal que ocupaba el volumen del tránsito por horas, siendo el logaritmo de $1 = 0$ correspondiente para el máximo volumen de peatones de todo el año en esa hora, $\log(3.596.022) = 15,09$ que corresponde con el acumulado a las 20:00 horas, seguido de las 21:00 horas. Las horas con menos tránsito fueron las 6:00 y las 5:00 horas respectivamente.

Gráficamente observamos que los datos presentan una tendencia lineal con pendiente negativa, lo que es congruente con una Ley de Potencias, aunque se confirmará a continuación con el enfoque analítico ajustando los puntos a una recta. El análisis de la esfericidad de los residuos se muestra en la Figura 1 del Anexo.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    14.1675    1.4006  10.115 9.79e-10 ***
tramo_1[c(1:24), 2] -0.8542    0.1005  -8.501 2.12e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4102 on 22 degrees of freedom
Multiple R-squared:  0.7666,    Adjusted R-squared:  0.756
F-statistic: 72.26 on 1 and 22 DF,  p-value: 2.123e-08

```

Figura 5.2.2. Salida del análisis de regresión tamaño-rango para la variable logaritmo del rango (x) frente al logaritmo del tamaño, número de peatones en el año 2020. Fuente: Elaboración propia.

En la figura 5.2.2 vemos que el modelo es estadísticamente significativo ya que el p valor del estadístico de contraste global del modelo es un $p < 0.05$ y el coeficiente de determinación ajustado R^2 es de 0.756 lo que quiere decir que la recta que ajustamos explica el 75,6% de la varianza de los datos proyectados.

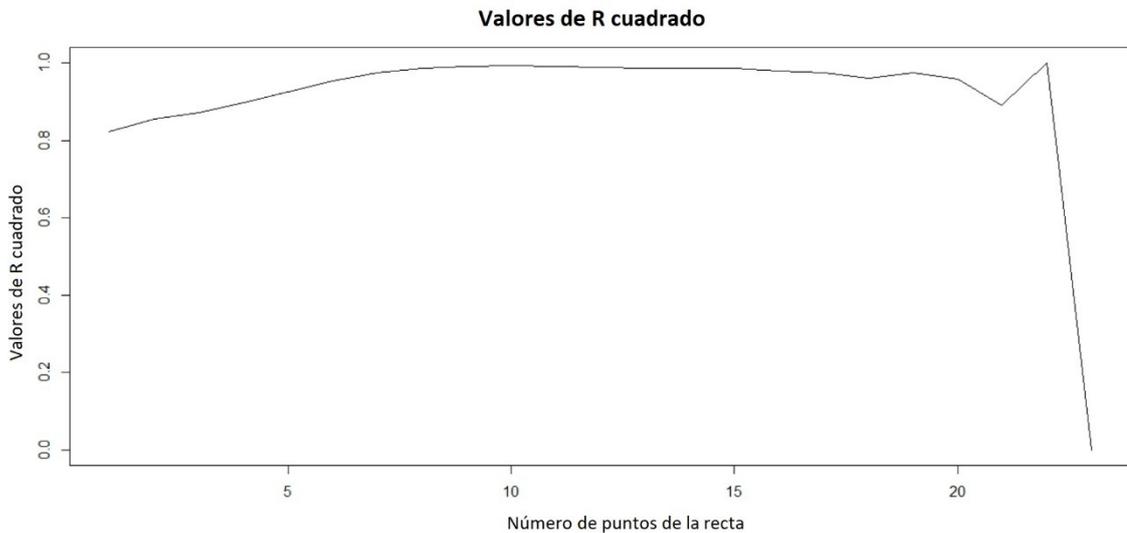


Figura 5.2.3. Representación de los valores de R^2 para el período 2020. Fuente: Elaboración propia

Una nota a tener en cuenta es que esta distribución no es necesario que se cumpla en todo el conjunto de datos, sino que donde verdaderamente tiene valor encontrar la validez es en la parte alta de la distribución por lo que empleamos la técnica de Clauset et al (2009). Esta técnica consiste en ir eliminando el punto anterior y generando una

nueva regresión con los puntos restantes (eliminando el punto mayor a continuación los dos puntos mayores y así sucesivamente...) y de este modo hasta que encontremos un máximo relativo en el gráfico de coeficientes de determinación, como vemos en la figura 5.2.3.

En el gráfico 5.2.3, se muestra la gráfica de los coeficientes R^2 prescindiendo de las observaciones sucesivas que se han eliminado, como se señaló en la metodología al principio de este capítulo. Se tomó la regresión a partir del décimo punto, donde vemos un máximo local hasta el final de la distribución. Todos los puntos ya cumplían unos coeficientes de determinación más que aceptables para poder afirmar que la distribución se cumplía en toda la muestra, pero siguiendo la metodología de Clauset et al. (2009) lo hemos dispuesto de esta forma.

No hay un consenso que determine cuantos puntos deben de quedar a la derecha del óptimo para que la regresión tenga valor. En este caso con tres puntos puede ajustarse satisfactoriamente, pero de no cumplirse un ajuste general que arroje un R^2 aceptable y un estadístico de contraste significativo en la regresión general, no tendría sentido buscar este óptimo del ajuste que permite ver la zona clave de los datos sobre los que se apoya la ley.

En el siguiente gráfico, figura 5.2.4 vemos el ajuste parcial a partir del décimo punto, donde se situaba el máximo local, este lo llevaremos a cabo por la metodología tradicional de MCO y por la planteada por Gabaix e Ibragimov. (2011).

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	8.5197	0.1449	58.82	< 2e-16 ***
tramo_1[c(10:24), 2]	-0.4269	0.0108	-39.54	6.23e-15 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 0.02602 on 13 degrees of freedom				
Multiple R-squared: 0.9918, Adjusted R-squared: 0.9911				
F-statistic: 1563 on 1 and 13 DF, p-value: 6.228e-15				

Figura 5.2.4. Ajuste del subconjunto por el método MCO. Fuente: Elaboración Propia

En la figura 5.2.4 ajustamos por MCO y los resultados presentan un ajuste excelente, con un estadístico F que abala el modelo ya que el p valor es inferior a 0,05. La pendiente negativa confirma el ajuste descendiente y el R^2 roza la unidad, no obstante, para muestras pequeñas aplicamos también el método de Gabaix que consiste en restar un medio al logaritmo del rango, $R(x_i)$.

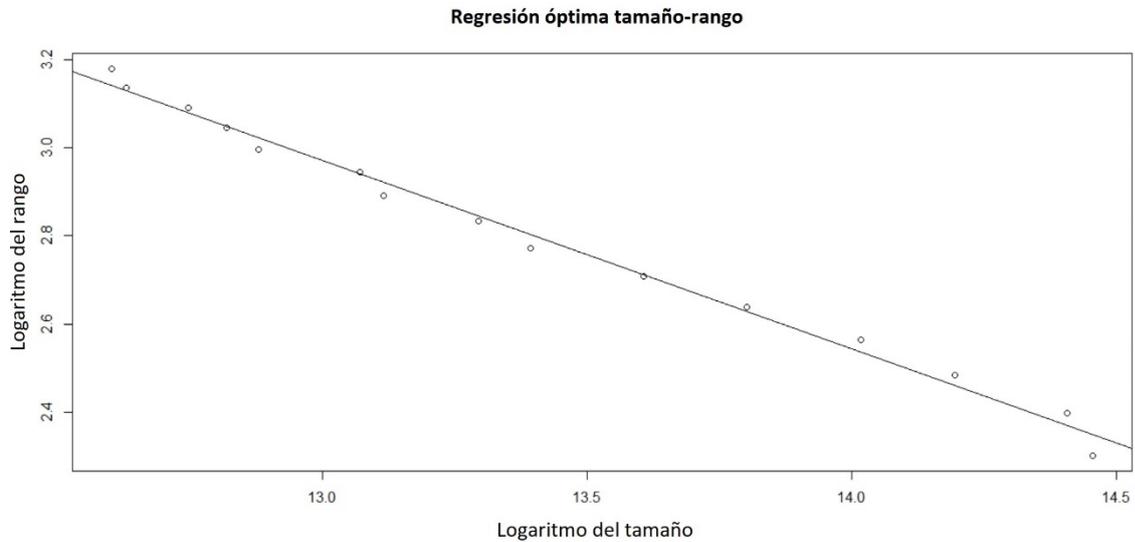


Figura 5.2.5. Regresión óptima tamaño-rango para el período 2020. Fuente: Elaboración Propia

En la figura 5.2.5 se observa la recta de regresión ajustada a los puntos del segmento que tomamos para verificar la ley de potencia de forma óptima.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.68008    0.14929   58.14 <2e-16 ***
datosx[c(10:24)] -0.44123    0.01113  -39.65  6e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02682 on 13 degrees of freedom
Multiple R-squared:  0.9918,    Adjusted R-squared:  0.9912
F-statistic: 1572 on 1 and 13 DF,  p-value: 6e-15

```

Figura 5.2.6. Ajuste del subconjunto por el método de Gabaix e Ibragimov para el período 2020. Fuente: Elaboración Propia

En la figura 5.2.6, se muestra la salida que arroja el método de Gabaix e Ibragimov, de modo que se mejora el R^2 asumiendo la misma validez en los estadísticos globales del modelo. Se afirma que el modelo es robusto frente al ajuste de Gabaix ya que no hay mucha diferencia entre los dos modelos. Al ser una muestra pequeña se elige este último.

4.1.6 5.1.2 Ajuste del período de confinamiento

Como se comentó al comienzo de este capítulo en los subperíodos analizados pretendemos encontrar una demostración similar en el periodo de confinamiento que se dio en la primavera de 2020, entre el 15 de marzo y el 15 de mayo. Durante este segmento temporal la movilidad sufrió un ligero cambio ya que en las dos horas más transitadas eran las 23:00 y las 22:00. En el siguiente gráfico, análogo al del apartado 5.1.1, se muestran datos del período de confinamiento.

El gráfico 5.2.7 muestra la acumulación del volumen de peatones por horas estudiada en escala log-log. La gráfica ja sido obtenida de manera análoga a la de la figura 5.2.1. que era para el período completo, el año 2020. En el gráfico vemos que para

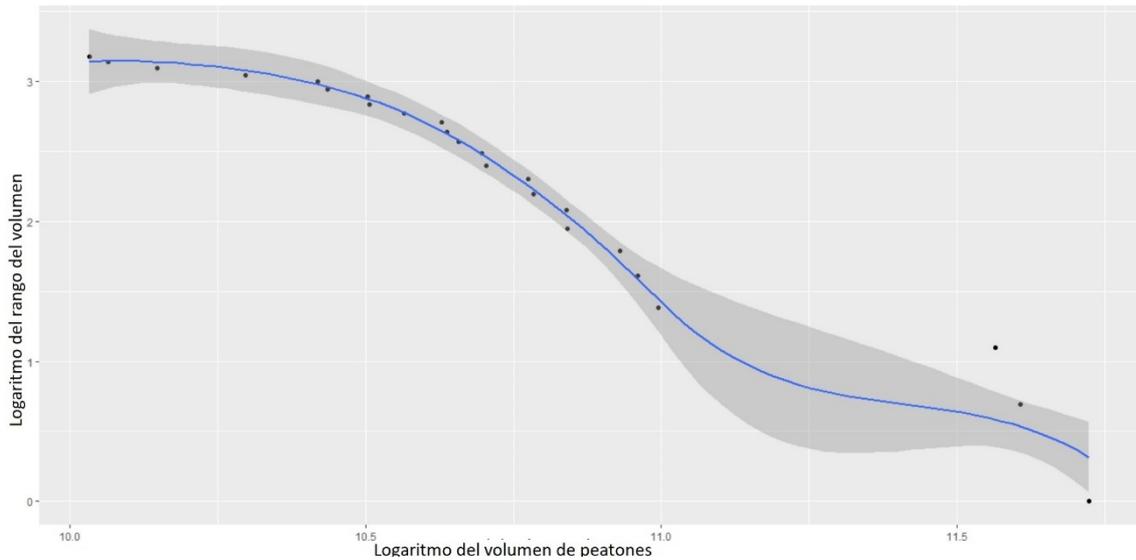


Figura 5.2.7. Representación del logaritmo del volumen de peatones y el del rango para el período de confinamiento. Fuente: Elaboración propia

este subperíodo tenemos una pendiente descendente sobre la cual podemos ajustar una recta de manera bastante precisa, lo que se determinará a continuación. En el siguiente análisis confirmamos la presencia de esta ley. El análisis de la esfericidad de los residuos se muestra en la Figura 2 del Anexo.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      21.7704     1.1997   18.15 1.01e-14 ***
tramo_2[c(1:24), 2] -1.8176     0.1118  -16.26 9.63e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2354 on 22 degrees of freedom
Multiple R-squared:  0.9232,    Adjusted R-squared:  0.9197
F-statistic: 264.3 on 1 and 22 DF,  p-value: 9.632e-14

```

Figura 5.2.8. Salida del análisis de regresión tamaño-rango para la variable logaritmo del rango (x) frente al logaritmo del tamaño, número de peatones en el período de confinamiento. Fuente: Elaboración propia

En la figura 5.2.8, se observa que el ajuste de los puntos tiene un coeficiente R^2 de 92,31 lo cual explica la mayoría de la variabilidad de los datos, el estadístico de validación global es significativo ya que se rechaza la hipótesis nula con el p valor < 0,05. Se parte de que todo este subperíodo del año 2020 cumple la ley de potencias al tener un ajuste bueno. De cualquier manera, siguiendo la metodología de Clauset et al. (2009) se va a determinar en que segmento se ajusta mejor.

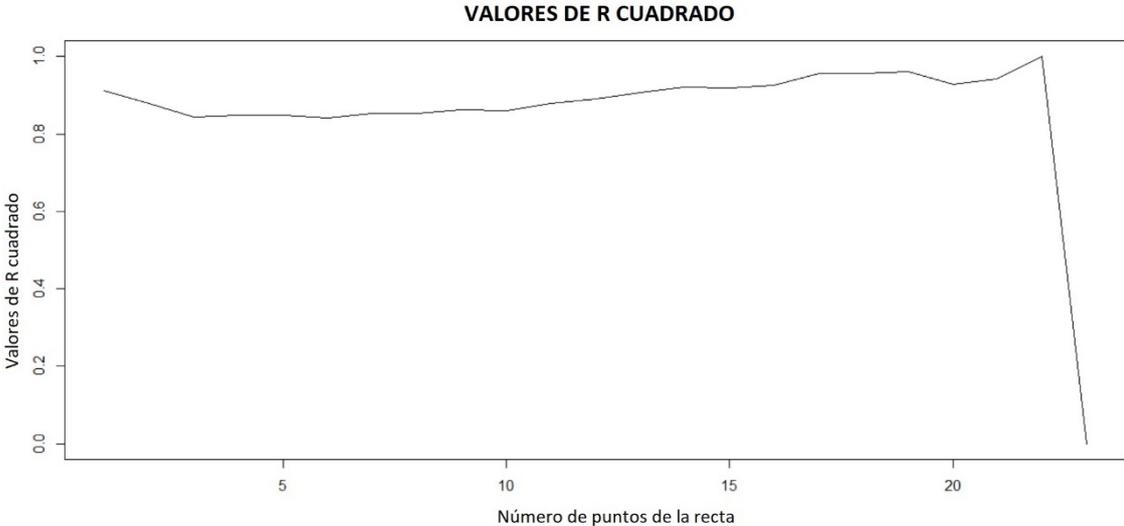


Figura 5.2.9 Representación de los valores de R^2 para el período de confinamiento. Fuente: Elaboración propia

Como se ve en la figura 5.2.9, se parte desde el punto 18 para hacer un segundo ajuste por las dos metodologías explicadas, el segundo ajuste confirma que la recta ajusta muy bien los puntos, el coeficiente de determinación es próximo a la unidad.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.4997     0.5209   16.32 1.58e-05 ***
tramo_2[c(18:24), 2] -0.5315     0.0507  -10.48 0.000136 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02365 on 5 degrees of freedom
Multiple R-squared:  0.9565,    Adjusted R-squared:  0.9478
F-statistic: 109.9 on 1 and 5 DF,  p-value: 0.0001363

```

Figura 5.2.10. Ajuste del subconjunto del confinamiento por el método MCO. Fuente: Elaboración Propia

En el ajuste anterior, correspondiente a la figura 5.2.10, se aprecia que el ajuste es preciso, con un p valor inferior a 0,05 y se asume que el modelo es globalmente significativo según el estadístico F. A continuación, en la figura 5.2.11 se aprecia la recta de regresión óptima y concuerda con el R^2 que se estimó.

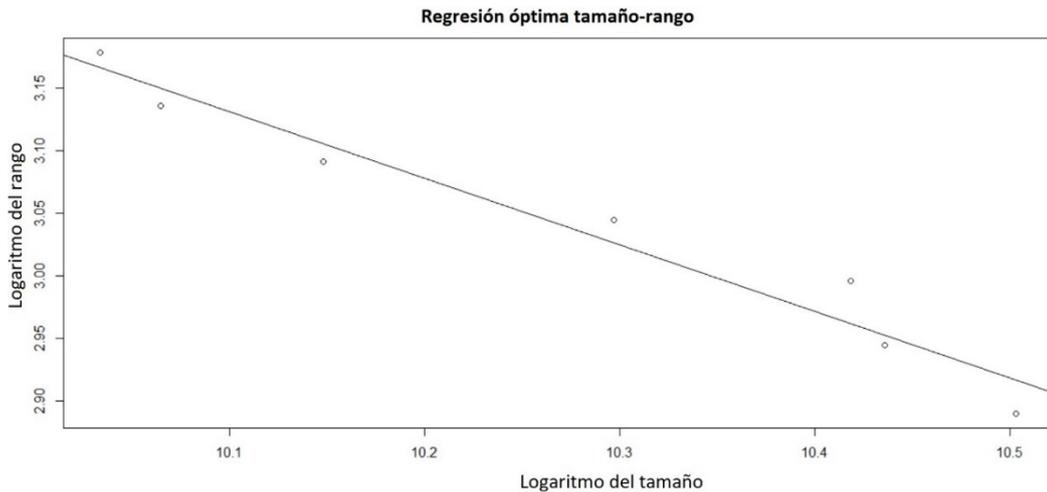


Figura 5.2.11. Regresión óptima tamaño-rango para el periodo de confinamiento. Fuente: Elaboración Propia

De manera análoga que en los apartados anteriores vamos a estimar la regresión propuesta por Gabaix que resulta como complemento al análisis para conjuntos de datos más pequeños. En este caso el coeficiente de determinación es similar al anterior. Por ende, se afirma la presencia de la ley de potencias a pesar de que esta segunda no mejora de manera sustancial el ajuste estándar por MCO.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.60970    0.53503   16.09 1.69e-05 ***
datosx_conf[c(18:24)] -0.54462    0.05208  -10.46 0.000138 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0243 on 5 degrees of freedom
Multiple R-squared:  0.9563,    Adjusted R-squared:  0.9475
F-statistic: 109.4 on 1 and 5 DF,  p-value: 0.0001379

```

Figura 5.2.12. Ajuste del subconjunto por el método de Gabaix e Ibragimov para el período de confinamiento. Fuente: Elaboración Propia

Como se observa en la figura 5.2.12 la regresión ajusta con un coeficiente de determinación de 0,956 y un estadístico de contraste F que otorga una validez global al modelo ya que el p valor es < 0,05.

Con este contraste tanto gráfico como analítico por las dos vías se afirma el cumplimiento de esta ley para el segmento temporal marzo-mayo. A priori se partía de que los datos podrían no ajustarse a este modelo debido a que descendió el volumen de movilidad significativamente. Sin embargo, el contraste ha verificado el cumplimiento de esta ley con esta casuística determinada para el distrito Centro de la ciudad de Madrid durante este periodo.

4.1.7 5.1.3 Ajuste del período estival.

Otra de las subdivisiones que se estableció en los objetivos fue la de analizar estos mismos datos para el periodo de verano en la ciudad de Madrid. Este período no hace referencia al verano como estación sino a los meses de julio y agosto. Madrid se caracteriza por tener un descenso importante de movilidad en el mes de agosto.

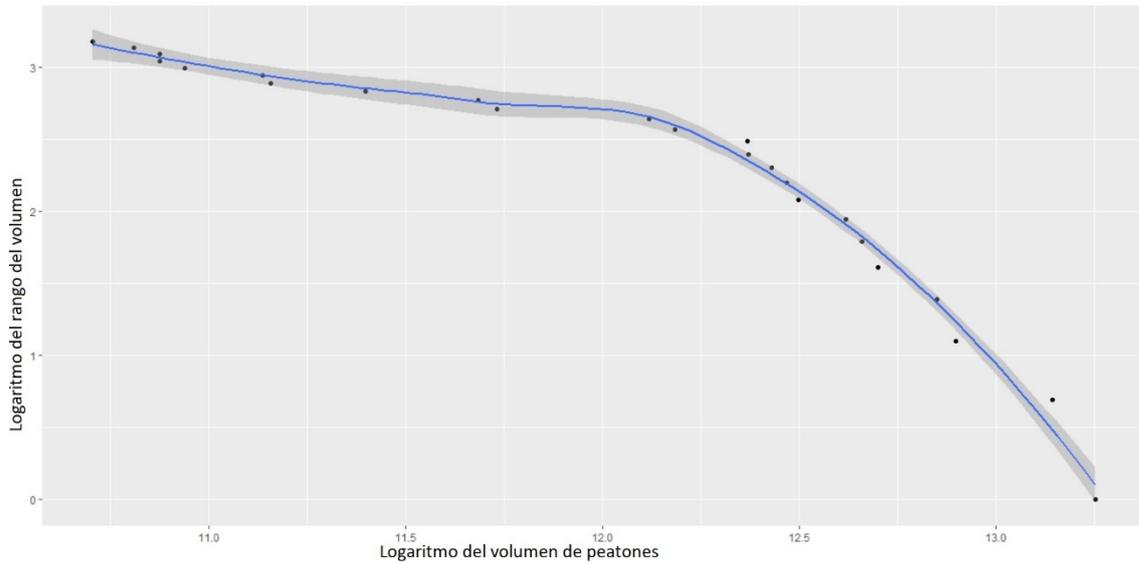


Figura 5.2.13 Representación del logaritmo del volumen de peatones y del rango para el periodo estival, julio - agosto 2020. Fuente: Elaboración propia

En el gráfico anterior, figura 5.2.13 se observa el volumen de peatones acumulado por horas, representado, como se menciona en la metodología en un sistema doblemente logarítmico. En este período estival la movilidad se ha caracterizado por tener sus horas punta a las 21:00 h y a las 20:0 h. Las horas con menos movilidad fueron las 5:00 y las 6:00 h. En la gráfica anterior se observa que la pendiente es negativa y que este segmento temporal puede ajustarse a una *Power Law* lo que contrastaremos de forma analítica a continuación.

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13.0237	1.2220	10.657	3.74e-10 ***
tramo_3[c(1:24), 2]	-0.8955	0.1017	-8.809	1.15e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 0.3991 on 22 degrees of freedom				
Multiple R-squared: 0.7791, Adjusted R-squared: 0.7691				
F-statistic: 77.6 on 1 and 22 DF, p-value: 1.15e-08				

Figura 5.1.14. Salida del análisis de regresión tamaño-rango para la variable logaritmo del rango (x) frente al logaritmo del tamaño, número de peatones del período estival. Fuente: Elaboración propia

En la figura 5.2.14 se observa que el ajuste para el período es bueno, siendo significativo el contraste global del modelo y con un R^2 de 0,77. El coeficiente de determinación no es lo suficientemente alto como para confirmar la presencia de una Power Law inmediatamente, por lo que se buscará el máximo local de los valores de R^2 aplicando la metodología de Clauset. La validación normalidad de los residuos se presenta en la Figura 3 del Anexo.

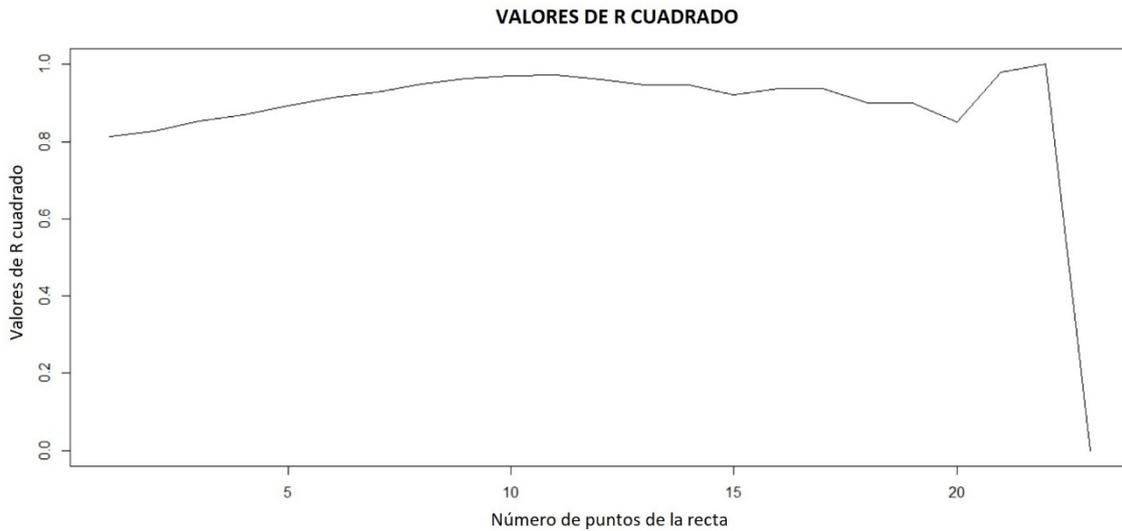


Figura 5.2.15. Representación de los valores de R^2 para el período estival. Fuente: Elaboración propia

Como se aprecia en la figura 5.2.15, se parte desde el punto 11 para hacer un segundo ajuste por las dos metodologías explicadas. El segundo ajuste confirma que la recta se ajusta correctamente, el coeficiente de determinación es próximo a la unidad con un coeficiente de 0,97 y un estadístico significativo como se ve a continuación en la figura 5.2.16

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      7.4109    0.2294   32.30 4.89e-13 ***
tramo_3[c(11:24), 2] -0.3996    0.0200  -19.98 1.41e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04421 on 12 degrees of freedom
Multiple R-squared:  0.9708,    Adjusted R-squared:  0.9684
F-statistic: 399 on 1 and 12 DF,  p-value: 1.415e-10

```

Figura 5.2.16. Ajuste del subconjunto del período estival por el método MCO. Fuente: Elaboración Propia

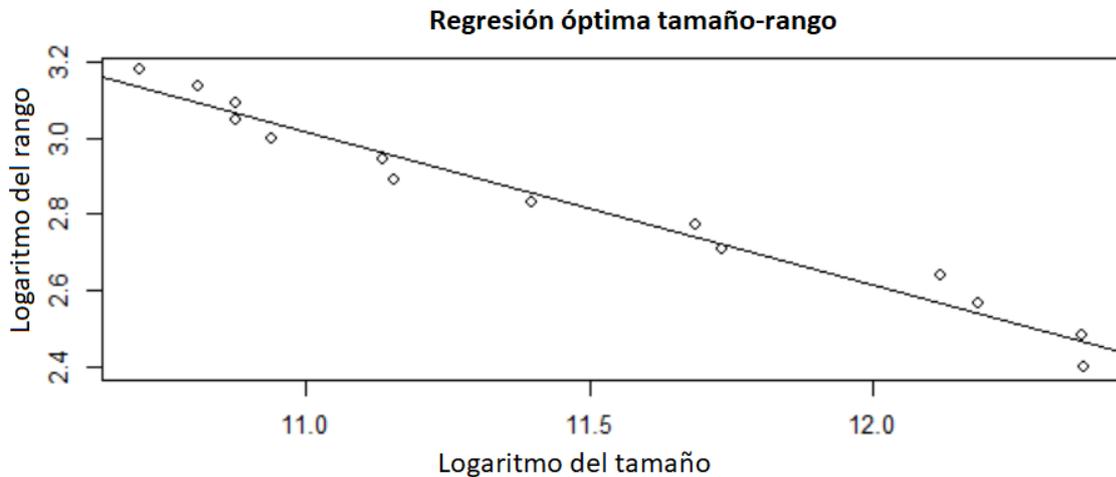


Figura 5.2.17. Regresión óptima tamaño-rango para el período estival. Fuente: Elaboración Propia

En la figura 5.2.17 se aprecia que el ajuste de los puntos con la recta de regresión a partir del óptimo es razonable. Ahora se compara con la metodología de Gabaix e Ibragimov. De cualquier modo, se puede afirmar que la ley de potencias está presente en este segmento.

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.52659	0.23662	31.81	5.88e-13 ***
datosx_verano[c(11:24)]	-0.41235	0.02063	-19.99	1.41e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 0.0456 on 12 degrees of freedom				
Multiple R-squared: 0.9708, Adjusted R-squared: 0.9684				
F-statistic: 399.5 on 1 and 12 DF, p-value: 1.406e-10				

Figura 5.2.18. Ajuste del subconjunto por el método de Gabaix e Ibragimov para el período de confinamiento. Fuente: Elaboración Propia

Como se observa en la figura 5.2.18 la regresión ajusta con un coeficiente de determinación de 0,97 y un estadístico de contraste F que otorga una validez global al modelo ya que el p valor es < 0,05.

Con este contraste tanto gráfico como analítico por las dos vías se afirma el cumplimiento de esta ley para el segmento temporal de julio y agosto en el cual la movilidad suele descender con respecto al resto del año. En el año 2020 este suceso ha sido al revés como se ve en la figura 5.1.1. Este período fue de recuperación de la

movilidad ya que se venía del descenso atípico generado por el confinamiento. Tras ejecutar la metodología se puede afirmar que sigue una *Power Law*

Los tres escenarios planteados en un primer momento siguen leyes de potencia. Esta conclusión que se ha obtenido para los datos del distrito Centro puede ser extrapolada a otros distritos y ciudades ya que los patrones de movilidad peatonal suelen ser los mismos.

5. CONCLUSIONES Y TRABAJO FUTURO

Las conclusiones del presente trabajo son las siguientes:

- Los datos de la movilidad de peatones por hora del distrito Centro de la ciudad de Madrid, en los tres períodos temporales considerados, siguen *leyes de potencia* o *power law*. Esta conclusión se ha obtenido a partir de las metodologías propuestas por Clauset et al. (2009) y Gabaix e Ibraginov (2011).
- Se ha contrastado y validado una *ley universal* sobre la cual se puede afirmar que la movilidad peatonal acumulada por segmentos temporales, en este caso horas, *sigue una distribución de ley de potencia*. La ley de potencias se ha cumplido en tres periodos diferentes de tiempo.
- Se he obtenido una *base de datos manejable*, relativa a movilidad de peatones por horas mediante sensores haciendo de la base de datos abierta del Ayuntamiento de Madrid. La metodología aplicada se basa en las técnicas de análisis de datos desarrolladas durante el Máster.
- Durante la recogida de los datos se encontró un error de medición de los sensores sobre el cual se informó al Ayuntamiento de Madrid y ya está corregido. Los sensores proporcionaban datos al *Datawarehouse* con un carácter en blanco el cual se almacenaba como “Centro” en vez de “Centro”. La base de datos trabajada con ese error proporcionaba 80.000 registros menos de los que necesitábamos solamente nos llegaban datos de las mediciones de calle de Fuencarral y Calle de San Bernardo.
- Partiendo de que se conoce la distribución estadística del volumen acumulado por horas de los peatones en un determinado punto, se pueden llevar a cabo cálculos de la probabilidad que existe de que haya x número de peatones en un determinado momento y utilizar esa estimación con distintos fines. Por ejemplo, se pueden aplicar la seguridad vial, o el volumen de personas que pasarán por el punto a en un rango horario con fines comerciales, entre otros.
- Como trabajo futuro, la misma ley de potencias se pueden contrastar para el caso de otros distritos y otras ciudades, ya que los patrones de movilidad peatonal suelen ser los mismos.

6. REFERENCIAS

- BALTHROP, Andrew; QUAN, Siyu. The power-law distribution of cumulative coal production. *Physica A: Statistical Mechanics and its Applications*, 2019, vol. 530, p. 121573.
- CLAUSET, Aaron; SHALIZI, Cosma Rohilla; NEWMAN, Mark EJ. Power-law distributions in empirical data. *SIAM review*, 2009, vol. 51, no 4, p. 661-703.
- GABAIX, Xavier; IBRAGIMOV, Rustam. Rank- $1/2$: a simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics*, 2011, vol. 29, no 1, p. 24-39.
- GABAIX, Xavier. Power laws in economics: An introduction. *Journal of Economic Perspectives*, 2016, vol. 30, no 1, p. 185-206
- KLAUS, Andreas; YU, Shan; PLENZ, Dietmar. Statistical analyses support power law distributions found in neuronal avalanches. *PloS one*, 2011, vol. 6, no 5, p. e19779.
- LORENZO, M. González; FERNÁNDEZ, María Lameiras; LOVELLE, MA Varela. Escalamiento de magnitud en la satisfacción laboral: I. Validez de la ley de potencia. *Revista de psicología general y aplicada: Revista de la Federación Española de Asociaciones de Psicología*, 1990, vol. 43, no 3, p. 411-417.
- XIN, Xiuying, et al. Power-law in pedestrian crossing flow under the interference of vehicles at an un-signalized midblock crosswalk. *Physica A: Statistical Mechanics and its Applications*, 2014, vol. 406, p. 287-297.
- XU, Xiao-Ting, et al. Understanding the diversity on power-law-like degree distribution in social networks. *Physica A: Statistical Mechanics and its Applications*, 2019, vol. 525, p. 576-581.
- YAMAMOTO, Keizo; MIYAZIMA, Sasuke. Power-law behavior in social and economical phenomena. *Physica A: Statistical Mechanics and its Applications*, 2004, vol. 344, no 3-4, p. 757-763.

Ticpymes. *Ayto. de Madrid implanta sistema de movilidad inteligente*. [en línea] disponible en <<https://www.ticpymes.es/tecnologia/noticias/1101343049504/ayuntamiento-de-madrid-implanta-sistema-de-movilidad-inteligente.1.html> 3.1.1> [julio 2018]

Yanner Barmm-Yam. *Power Law Concepts*. [en línea] disponible en <<https://necsi.edu/power-law>> [agosto de 2021]

Portal de datos abiertos del Ayuntamiento de Madrid. *Aforos de peatones y bicicletas* [en línea] disponible en <<https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=695cd64d6f9b9610VgnVCM1000001d4a900aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnnextfmt=default>> [abril 2021]

Portal de datos abiertos del Ayuntamiento de Madrid (2021) *Fichero de datos de movilidad* [en línea] disponible en <https://datos.madrid.es/FWProjects/egob/Catalogo/Transporte/Ficheros/Estructura_DS_aforos_peaton_bicis.pdf> [20 abril 2021]

Instituto de Estadística de la Comunidad de Madrid [en línea] disponible en <<http://www.madrid.org/iestadis/fijas/estructu/demograficas/padron/estructupcpr.htm>> [mayo de 2021]

ANEXO

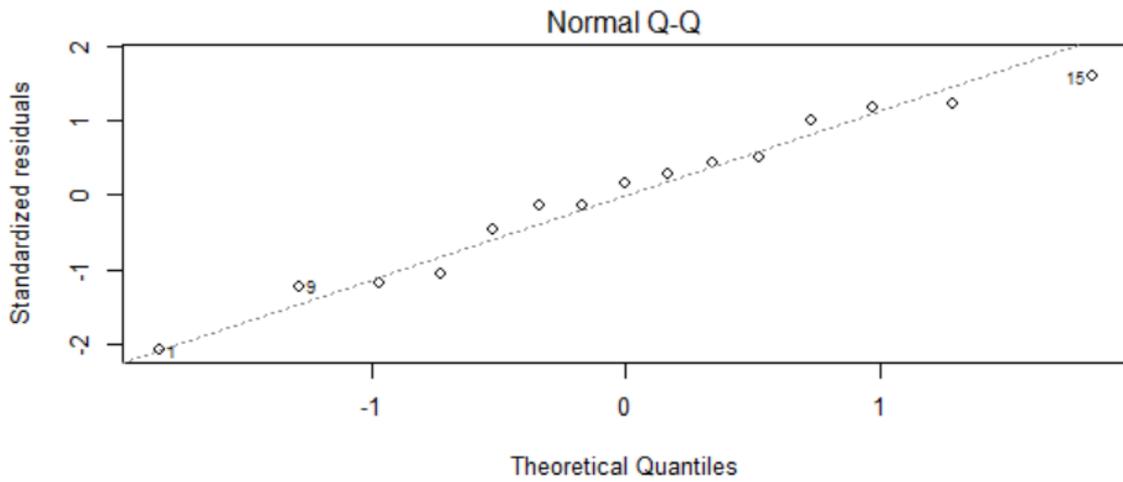


FIGURA 1. Gráficos de normalidad y residuos del periodo 2020.

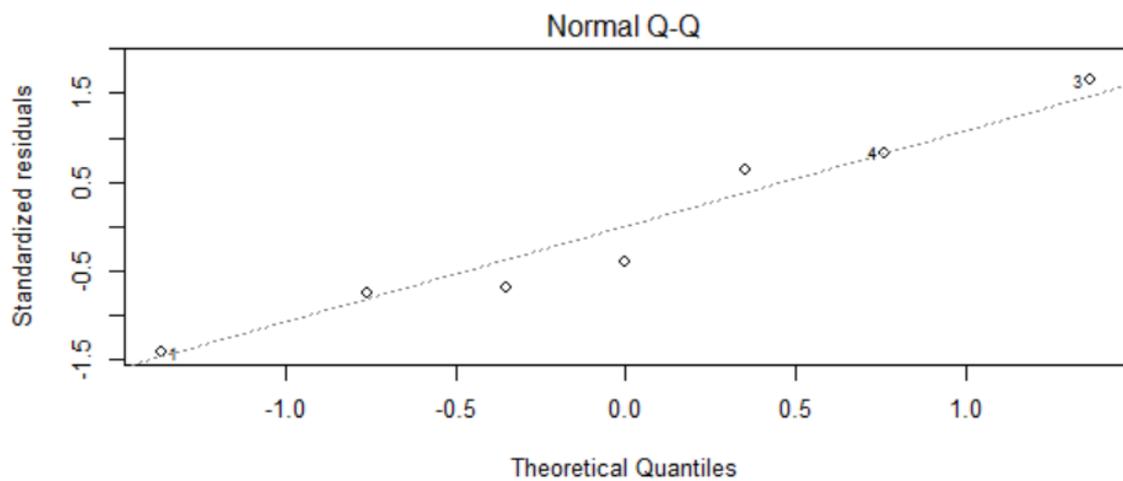


FIGURA 2. Gráficos de normalidad y residuos del periodo de confinamiento

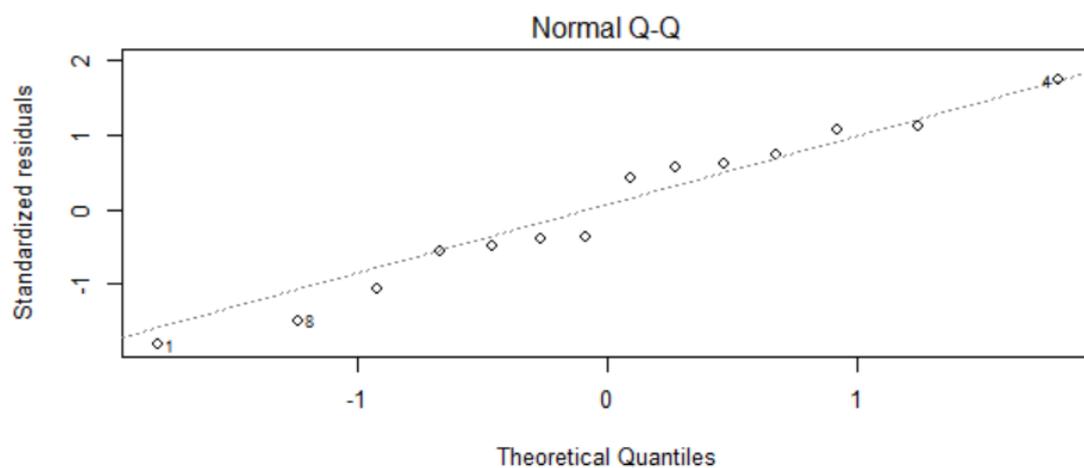


FIGURA 3. Gráficos de normalidad y residuos del período estival