



Facultad de Ciencias Económicas y Empresariales

Data mining para el estudio de los niveles de contaminación en el sexenio 2017-2022 en Canberra, Australia.

Autor: Carlos Pérez Olid

Director: Manuel Alejandro Betancourt Odio

MADRID | Junio 2023

RESUMEN / PALABRAS CLAVE

Australia es uno de los países con mejor calidad del aire del mundo. De hecho, en 2021 fue uno de los 13 países del mundo que cumplieron con los niveles de concentración de PM_{2,5} que recomienda la Organización Mundial de la salud, no obstante, es un país que afronta múltiples desafíos en cuanto contaminación atmosférica como consecuencia de los incendios forestales. En concreto, la capital de Australia, Canberra, que es el medio objeto de investigación de este trabajo destaca por ser una ciudad con una topografía particular. Asimismo, Australia implementó un confinamiento muy restrictivo durante la pandemia a raíz del COVID-19. Por tanto, en este trabajo se ha realizado un estudio de la evolución de los niveles de contaminación de NO₂, O₃ y PM₁₀ en el sexenio 2017-2022 en Canberra. Por un lado, se ha hecho especial hincapié en el impacto que tuvo en los niveles de contaminación la temporada de incendios 2019-2020 puesto que fue una de las más destructivas de la historia del país. Por otro lado, las restricciones de movilidad que se implementaron como consecuencia del COVID-19 proporcionan una oportunidad única para comprender el impacto que tiene la actividad humana sobre la contaminación.

Para llevar a cabo el estudio se han tenido en cuenta las variables de climatología que afectan a la contaminación (temperatura y precipitaciones) y se ha elaborado un modelo de regresión lineal múltiple. El objetivo final del estudio es determinar que variables afectan a los niveles de contaminación, sus causas y sus patrones de comportamiento.

Los resultados muestran, por un lado, que el confinamiento por el COVID-19 tuvo un efecto heterogéneo entre los contaminantes analizados y, por otro lado, que los incendios de la temporada 2019-2020 tuvieron un efecto devastador en los niveles de calidad del aire en Canberra.

Palabras clave: Canberra, Contaminación, COVID-19, Incendios forestales, Efecto fin de semana, Dióxido de nitrógeno, Ozono, Partículas volátiles.

ABSTRACT / KEY WORDS

Australia is one of the countries with the best air quality in the world. In fact, in 2021 it was one of the 13 countries in the world that met the $PM_{2.5}$ concentration levels recommended by the World Health Organization. However, it is a country that faces multiple challenges regarding air pollution as a result of wildfires. Specifically, the capital of Australia, Canberra, which is the focus of this study, stands out as a city with a unique topography. Additionally, Australia implemented a very strict lockdown during the COVID-19 pandemic. Therefore, this study has examined the evolution of NO_2 , O_3 , and PM_{10} pollution levels in Canberra from 2017 to 2022. On one hand, special emphasis has been placed on the impact of the 2019-2020 wildfire season, as it was one of the most destructive in the country's history. On the other hand, mobility restrictions implemented due to COVID-19 provide a unique opportunity to understand the impact of human activity on pollution.

The study takes into account climatic variables that affect pollution (temperature and precipitations) and has developed a multiple linear regression model. The ultimate goal of the study is to determine which variables affect pollution levels, their causes, and their behavioral patterns.

The results show, on one hand, that the COVID-19 lockdown had a heterogeneous effect on the analyzed pollutants, and on the other hand, that the 2019-2020 wildfires had a devastating effect on air quality levels in Canberra.

Key words: Canberra, Pollution, COVID-19, Wildfires, Weekend effect, Nitrogen dioxide, Ozone, Volatile particles.

ÍNDICE

RESUMEN / PALABRAS CLAVE	2
ABSTRACT / KEY WORDS	3
ÍNDICE DE FIGURAS	5
ÍNDICE DE TABLAS	8
INTRODUCCIÓN	9
CAPÍTULO I: ACTUALIDAD DEL TEMA	13
CAPÍTULO II: METODOLOGÍA	17
1. MATERIALES	17
2. METODOLOGÍA PARA EL ANÁLISIS DE LA CONTAMINACIÓN	24
CAPÍTULO III: APLICACIÓN DE LA METODOLOGIA	26
1. DESCRIPCIÓN DEL MEDIO OBJETO DE INVESTIGACIÓN	26
2. CONJUNTO DE DATOS	28
3. TRATAMIENTO DE DATOS CON R.....	34
CAPÍTULO IV: RESULTADOS	35
1. CONSTRUCCIÓN Y RESULTADOS DE LOS MODELOS ECONOMÉTRICOS DE REGRESIÓN	35
1.1 Resultados de los modelos de regresión en Monash	37
1.2 Resultados de los modelos de regresión en Florey	52
1.3 Resultados de los modelos de regresión en Civic.....	68
2. DISCUSIÓN DE LOS RESULTADOS OBTENIDOS	77
CAPÍTULO V: CONCLUSIONES	82
BIBLIOGRAFÍA	86
ANEXOS	97

ÍNDICE DE FIGURAS

Figura 1. Gráfico de la distribución de los datos por cada partícula	37
Figura 2. Diagrama de cajas de la concentración promedio de NO ₂ versus las variables independientes de tipo factor en Monash	38
Figura 3. Correlación de la concentración promedio de NO ₂ versus la temperatura y la precipitación en Monash.....	39
Figura 4. Comprobación del supuesto de linealidad del modelo de regresión de NO ₂ en Monash	40
Figura 5. Comprobación del supuesto de normalidad del modelo de regresión de NO ₂ en Monash	41
Figura 6. Comprobación del supuesto de homocedasticidad del modelo de regresión de NO ₂ en Monash	41
Figura 7. Valores influyentes en el modelo de concentración promedio de NO ₂	42
Figura 8. Diagrama de cajas de la concentración promedio de O ₃ versus las variables independientes de tipo factor en Monash	44
Figura 9. Correlación de la concentración promedio de O ₃ versus la temperatura y la precipitación en Monash.....	44
Figura 10. Comprobación del supuesto de linealidad del modelo de regresión de O ₃ en Monash	45
Figura 11. Comprobación del supuesto de normalidad del modelo de regresión de O ₃ en Monash	45
Figura 12. Comprobación del supuesto de homocedasticidad del modelo de regresión de O ₃ en Monash	46
Figura 13. Valores influyentes en el modelo de concentración promedio de O ₃	46
Figura 14. Diagrama de cajas de la concentración promedio de PM ₁₀ versus las variables independientes de tipo factor en Monash	48
Figura 15. Correlación de la concentración promedio de PM ₁₀ versus la temperatura y la precipitación en Monash.....	49
Figura 16. Comprobación del supuesto de linealidad del modelo de regresión de PM ₁₀ en Monash	49
Figura 17. Comprobación del supuesto de normalidad del modelo de regresión de PM ₁₀ en Monash	50

Figura 18. Comprobación del supuesto de homocedasticidad del modelo de regresión de PM_{10} en Monash	51
Figura 19. Valores influyentes en el modelo de concentración promedio de PM_{10}	51
Figura 20. Diagrama de cajas de la concentración promedio de NO_2 versus las variables independientes de tipo factor en Florey	53
Figura 21. Correlación de la concentración promedio de NO_2 versus la temperatura y la precipitación en Florey	54
Figura 22. Comprobación del supuesto de linealidad del modelo de regresión de NO_2 en Florey	55
Figura 23. Comprobación del supuesto de normalidad del modelo de regresión de NO_2 en Florey	55
Figura 24. Comprobación del supuesto de homocedasticidad del modelo de regresión de NO_2 en Florey	56
Figura 25. Valores influyentes en el modelo de concentración promedio de NO_2	56
Figura 26. Diagrama de cajas de la concentración promedio de O_3 versus las variables independientes de tipo factor en Florey	58
Figura 27. Correlación de la concentración promedio de O_3 versus la temperatura y la precipitación en Florey	59
Figura 28. Comprobación del supuesto de linealidad del modelo de regresión de O_3 en Florey	60
Figura 29. Comprobación del supuesto de normalidad del modelo de regresión de O_3 en Florey	60
Figura 30. Comprobación del supuesto de homocedasticidad del modelo de regresión de O_3 en Florey	61
Figura 31. Valores influyentes en el modelo de concentración promedio de O_3	61
Figura 32. Diagrama de cajas de la concentración promedio de PM_{10} versus las variables independientes de tipo factor en Florey	63
Figura 33. Correlación de la concentración promedio de PM_{10} versus la temperatura y la precipitación en Florey	64
Figura 34. Comprobación del supuesto de linealidad del modelo de regresión de PM_{10} en Florey	65
Figura 35. Comprobación del supuesto de normalidad del modelo de regresión de PM_{10} en Florey	65
Figura 36. Comprobación del supuesto de homocedasticidad del modelo de regresión	66

Figura 37. Valores influyentes en el modelo de concentración promedio de PM_{10}	66
Figura 38. Diagrama de cajas de la concentración promedio de O_3 versus las variables independientes de tipo factor en Civic	68
Figura 39. Correlación de la concentración promedio de O_3 versus la temperatura y la precipitación en Civic	69
Figura 40. Comprobación del supuesto de linealidad del modelo de regresión de O_3 en Civic	69
Figura 41. Comprobación del supuesto de normalidad del modelo de regresión de O_3 en Civic	70
Figura 42. Comprobación del supuesto de homocedasticidad del modelo de regresión de O_3 en Civic	70
Figura 43. Valores influyentes en el modelo de concentración promedio de O_3	71
Figura 44. Diagrama de cajas de la concentración promedio de PM_{10} versus las variables independientes de tipo factor en Civic	73
Figura 45. Correlación de la concentración promedio de PM_{10} versus la temperatura y la precipitación en Civic	73
Figura 46. Comprobación del supuesto de linealidad del modelo de regresión de PM_{10} en Civic	74
Figura 47. Comprobación del supuesto de normalidad del modelo de regresión de PM_{10} en Civic.....	74
Figura 48. Comprobación del supuesto de homocedasticidad del modelo de regresión de PM_{10} en Civic	75
Figura 49. Valores influyentes en el modelo de concentración promedio de PM_{10}	75

ÍNDICE DE TABLAS

Tabla 1. Estaciones de monitorización de la calidad del aire en el ACT	19
Tabla 2. Estaciones de monitorización de la temperatura máxima en el ACT.....	21
Tabla 3. Estaciones de monitorización de las precipitaciones en el ACT	22
Tabla 4. Variables independientes utilizadas en los modelos de regresión.....	30
Tabla 5. Resultados del modelo de regresión para la partícula NO ₂ en Monash.....	43
Tabla 6. Resultados del modelo de regresión para la partícula O ₃ en Monash	47
Tabla 7. Resultados del modelo de regresión para la partícula PM ₁₀ en Monash	52
Tabla 8. Resultados del modelo de regresión para la partícula NO ₂ en Florey	57
Tabla 9. Resultados del modelo de regresión para la partícula O ₃ en Florey.....	62
Tabla 10. Resultados del modelo de regresión para la partícula PM ₁₀ en Florey.....	67
Tabla 11. Resultados del modelo de regresión para la partícula O ₃ en Civic.....	72
Tabla 12. Resultados del modelo de regresión para la partícula PM ₁₀ en Civic.....	76
Tabla 13. Resultados de la regresión por contaminante en Canberra.....	77

INTRODUCCIÓN

En la actualidad, una de las principales inquietudes de la sociedad radica en la protección y conservación del medio ambiente, con el propósito de reparar los perjuicios ocasionados causados por la humanidad y evitar que este siga deteriorándose.

Poniendo foco en la contaminación atmosférica, esta se presenta como uno de los desafíos medioambientales más críticos a nivel global. La contaminación atmosférica se encuentra presente en todas las sociedades, sin importar el grado de desarrollo socioeconómico, y destaca por su significativo impacto sobre la salud humana. Según la Organización Mundial de la Salud (2022): “Los efectos combinados de la contaminación del aire ambiente y la del aire doméstico se asocian a 6,7 millones de muertes prematuras cada año”. Asimismo, también afirman que “En 2019 la contaminación del aire ambiente (exterior) provocó en todo el mundo 4,2 millones de muertes prematuras”.

El presente proyecto se centra en el estudio de la contaminación atmosférica de la capital de Australia, Canberra. De hecho, Ember que es una organización especializada en el seguimiento de las emisiones de carbono, publicó en 2022 su informe Per Capita Coal Power Emissions en el que Australia encabezaba la lista de emisiones de CO₂ per cápita, triplicando la media global (Ember, 2022).

La contaminación atmosférica contribuye a diferentes problemas relacionados con la salud de las personas, incluyendo la mortalidad, el estrés psicológico y problemas cardiovasculares (Balasooriya et al., 2022). En el caso concreto de Australia, se estima que la exposición a la contaminación del aire provoca 4880 muertes prematuras cada año (Walter, 2019).

Canberra, la capital nacional de Australia y medio objeto de estudio de este proyecto, fue calificada durante 2019 como la tercera ciudad más contaminada de Australia y se ubicó como la 49^a capital más contaminada del mundo en términos de promedio anual de PM_{2,5} (partículas finas en suspensión de hasta 2,5 micrones de diámetro), dentro de un ranking de 84 capitales globales. Asimismo, es conocida como la "capital del bosque" del país, por sus extensas reservas naturales y las cadenas montañosas que la rodean, lo que provoca que en numerosas ocasiones el humo de los

incendios forestales llegue a la ciudad durante la temporada anual de incendios. Existen dos factores claves relacionados con la geografía de Canberra que pueden exacerbar el humo una vez que ha alcanzado la ciudad: su ubicación en un valle rodeado de montañas y sus patrones de viento. Por otro lado, es importante resaltar que la inversión térmica (muy presente en esta ciudad) atrapa la contaminación durante períodos más largos. (IQAir, s.f.)

Por último, las restricciones de movilidad, como consecuencia del COVID-19 han ofrecido una oportunidad única para estudiar los efectos de la actividad humana en la contaminación atmosférica. Los resultados de estos estudios pueden orientar a los gobiernos en relación con las medidas que han de adoptar para reducir la contaminación del aire. (Chossière et al., 2021)

Considerando los elementos hasta aquí expuestos se ha planteado el siguiente problema de investigación en el medio objeto de estudio:

Canberra se enfrenta a graves problemas de contaminación atmosférica y en ocasiones supera tanto los límites establecidos por las directivas de la Unión Europea como los estándares de calidad del aire recomendados por la Organización Mundial de la Salud (OMS) para proteger la salud de las personas.

La hipótesis que se plantea al problema de investigación es:

Durante el periodo de pandemia por la COVID-19, se implementaron diversas medidas de confinamiento que incluyeron restricciones a la movilidad y el distanciamiento social. Estas medidas tuvieron un impacto significativo en la reducción de los niveles de contaminación en Canberra y en la mejora de la calidad del aire. Además, es importante destacar que la contaminación en Canberra también se ve influenciada por la incidencia de incendios, ya que estos eventos pueden afectar considerablemente a la calidad del aire en la región.

A partir del planteamiento del problema, el objetivo general de la investigación es:

Realizar un análisis de la variación en la concentración de los principales contaminantes atmosféricos durante el periodo comprendido entre 2017 y 2022 teniendo en cuenta la pausa social que se implementó como consecuencia de la pandemia a raíz del COVID-19.

Para el logro del objetivo general planteado se hace necesario el planteamiento de los siguientes objetivos específicos:

- I. Identificar distintos patrones de comportamiento presentes en los índices representativos de la calidad del aire en función de las variables que influyen en la contaminación atmosférica, con el fin de establecer relaciones y correlaciones significativas.
- II. Examinar el comportamiento de diversos factores que influyen en la calidad del aire, más allá de las medidas implementadas durante la pandemia del COVID-19. Se prestará especial atención a variables como la climatología y los patrones de comportamiento que varían a lo largo de la semana.
- III. Estudiar las características del territorio de Canberra con el objetivo de determinar que influencia tienen estas sobre los niveles de contaminación como consecuencia de los grandes incendios de Australia.

Esta investigación está motivada por el interés en fortalecer las políticas orientadas a lograr una reducción efectiva de las emisiones y la contaminación derivadas de la actividad humana. La justificación y el análisis de su efectividad se respaldan en el confinamiento debido al COVID-19, el cual puede ser empleado como una prueba "experimental".

Dado lo expuesto anteriormente, el presente trabajo se ha articulado en cuatro capítulos, además de esta introducción en la que se expone la justificación del tema, el objeto de investigación, sus objetivos y la hipótesis de trabajo.

El primer capítulo se dedica un epígrafe a revisar la relevancia y actualidad del tema. Para ello, se ha realizado un estudio de los impactos y efectos de la contaminación

en la salud y en el medioambiente. Además de sus efectos se realiza un estudio de las fuentes de emisiones contaminantes, y de aquellos factores que pueden ser relevantes para el estudio de la calidad del aire.

En el segundo capítulo se expone la metodología y los materiales utilizados. Por un lado, se realiza una explicación de las distintas fuentes de datos que se han utilizado para el estudio de la contaminación en la ciudad de Canberra. En este capítulo también se presenta la herramienta que se ha utilizado para estudiar la contaminación, *RStudio* y se realiza una explicación del concepto y características de un modelo de regresión lineal.

El tercer capítulo está enfocado en la aplicación de la metodología. En un primer lugar se realiza una descripción del medio objeto de investigación y una explicación de las distintas variables que se han utilizado para realizar un modelo de regresión con el que se estudia la contaminación en la ciudad de Canberra. También se procede a realizar una explicación del proceso que se ha realizado para el tratamiento de los datos con la herramienta *RStudio*.

En el cuarto capítulo se presentan los resultados obtenidos de los modelos de regresión. Se ha realizado un modelo de regresión para cada una de las partículas contaminantes que se estudian en este trabajo (NO_2 , O_3 , PM_{10}) en cada una de las estaciones de medición de la ciudad de Canberra.

Finalmente, en el último capítulo se exponen las conclusiones obtenidas en la investigación.

CAPÍTULO I: ACTUALIDAD DEL TEMA

Como explica, Anwar et al. (2018), la lucha continua para mejorar los estilos de vida ha llevado al desarrollo continuo de la raza humana en las últimas décadas, sin embargo, el coste de este desarrollo se paga con los efectos adversos sobre el medio ambiente circundante (p. 1). Entre estos efectos adversos se encuentra la contaminación atmosférica. La contaminación atmosférica se ha definido como “la presencia en el aire de materia o formas de energía que impliquen riesgo, daño o molestia grave para las personas o bienes de cualquier naturaleza” (Martínez y Díaz, 2004, p. 13).

La contaminación atmosférica siempre ha existido. Las erupciones volcánicas o los incendios han sido elementos contaminantes, pero cuando el hombre descubre el fuego es cuando se introduce por primera vez la contaminación atmosférica antropogénica, es decir, la contaminación causada por el hombre. (Aránguez et al., 1999, p. 124). Sin embargo, las fuentes naturales de contaminación no son un fuerte motivo de preocupación, ya que forman parte del equilibrio del entorno natural mientras que la contaminación química antropogénica es motivo de gran preocupación ya que sus fuentes aumentan en número y concentración con el aumento de la población humana mundial y el continuo aumento de la demanda de energía (Kumar, 2010, p. 2). Ejemplos de fuentes de contaminación antropogénica son el uso de combustibles fósiles, los procesos industriales y el uso de disolventes, la agricultura y el tratamiento de residuos (Agencia Europea del Medio Ambiente, 2020, p. 3).

La contaminación atmosférica se encuentra entre los principales retos de la era contemporánea. No solo existe la amenaza del cambio climático y sus consecuencias para el hombre, sino que la contaminación atmosférica también supone una amenaza para la salud ya que se ha visto asociada con un aumento de la morbilidad y mortalidad. (Manisalidis et al., 2020, p. 1).

Según las estadísticas de la Organización Mundial de la Salud (OMS), el 99% de la población mundial respira niveles peligrosos de PM y NO₂, y las personas de países de ingresos bajos están expuestas a niveles de contaminación del aire que superan los estándares de calidad del aire que establece la OMS. Además, la OMS estima que la exposición a las partículas PM_{2,5} ha causado 6,7 millones de fallecimientos.

(Organización Mundial de la Salud, 2022, p. 1). Comúnmente se reconoce que la contaminación del aire es dañina para los pulmones y las vías respiratorias, pero también puede afectar a la mayoría de los demás sistemas de órganos del cuerpo humano (Schraufnagel et al., 2019).

Se estima que la contaminación ha causado alrededor de 500.000 muertes por cáncer de pulmón y 1,6 millones de muertes por obstrucción pulmonar, de hecho, múltiples estudios han demostrado que los riesgos también se extienden a enfermedades cardiovasculares, diabetes, obesidad, mortalidad prenatal y a trastornos neurológicos y del sistema inmunitario (Schraufnagel et al., 2019; Yadav, 2021). En Europa, la Agencia Europea del Medio Ambiente (AEMA) estima que, en 2020, aproximadamente 238.000 muertes prematuras fueron atribuibles al PM_{2,5} en los 27 estados miembros de la Unión Europea (AEMA, 2023). En 2018, un 2% de las muertes y el 1,3 % de la totalidad de enfermedades en Australia se atribuyeron a la contaminación del aire por PM_{2,5} (Instituto Australiano de Salud y Bienestar, 2022).

Además de los riesgos para la salud, la contaminación atmosférica agrava el cambio climático que amenaza el modo de vida de la sociedad. Las temperaturas globales del planeta están aumentando debido al aumento de los gases de efecto invernadero en la atmósfera (D'Amato, 2020). Los gases de efecto invernadero son necesarios para mantener la temperatura de la tierra, pero la emisión antropogénica de esos gases ha causado el calentamiento global (Shahzad, 2015). De acuerdo con Beniston (2021), entre las consecuencias del cambio climático se encuentran el aumento de las temperaturas que provoca el deshielo de los glaciares y un consecuente aumento del nivel del mar, lo cual pone en peligro el hogar de millones de personas que viven en zonas costeras (p. 562). Además, debido al cambio climático, se aumenta la frecuencia y gravedad de eventos climáticos extremos. Los aumentos de temperatura y los eventos climáticos extremos ponen en peligro la seguridad alimentaria, el acceso a agua potable y el bienestar de los humanos al fomentar la aparición y expansión de enfermedades como la malaria (Beniston, 2010, pp. 563-566).

A pesar de que las emisiones producidas por eventos naturales pueden tener efectos en el cambio climático, ello no explica el gran aumento de las temperaturas que se ha observado desde principios del siglo XX. Numerosos estudios apuntan a que el

cambio climático se debe a la actividad humana ya que desde la Revolución Industrial se han emitido grandes cantidades de gases de efecto invernadero a la atmósfera (Grupo Intergubernamental de Expertos sobre el Cambio Climático [IPCC], 2021, p. 4). Se ha estimado que alrededor de dos tercios de las emisiones de gases de efecto invernadero provienen de la combustión de combustibles que principalmente se desarrolla en el sector de la industria. El sector de la energía y el transporte también han sido identificados como los grandes contaminantes (Agencia de Protección Ambiental, 2023).

Al margen de la actividad industrial, es necesario tener en cuenta que el estudio de los niveles de contaminación en un determinado periodo se verá afectado tanto por las emisiones antropogénicas de contaminantes como por otros factores que afectarán a los resultados obtenidos y su interpretación. Estos factores son emisiones naturales como son los incendios, la meteorología y las medidas adoptadas para frenar la expansión del COVID-19.

Los incendios forestales liberan partículas nocivas conocidas como $PM_{2,5}$ y PM_{10} . Las $PM_{2,5}$, son tan pequeñas que pueden viajar profundamente a los pulmones, causando problemas de salud a corto plazo y agravando trastornos respiratorios a largo plazo (Kenward et al., 2013, p. 9). Se espera que en los próximos años los grandes eventos de contaminación debido a la quema de biomasa sean más recurrentes a medida que el cambio climático avanza, debido a las altas concentraciones de gases de efecto invernadero (Bolaño et al. 2022, p. 2). Además, las masas de aire con los contaminantes de la quema de biomasa se pueden transportar largas distancias, afectando a la calidad del aire de multitud de poblaciones alejadas del foco de incendio (Bolaño et al., 2022, p. 2).

Por otro lado, la meteorología, es decir, factores como la temperatura, precipitaciones y viento afectan a la concentración de contaminantes en la atmósfera. Estos factores tienen un gran impacto en la difusión, acumulación y transporte de contaminantes atmosféricos, y provocan grandes cambios en los niveles de contaminación (Feng et al., 2012). Estudios han demostrado que las precipitaciones pueden mejorar efectivamente la calidad del aire porque el agua de lluvia puede eliminar $PM_{2,5}$ y otros contaminantes (Wang, 2023, p. 1). En cuanto a las altas temperaturas, cuando estas se dan en ambientes secos, contribuyen significativamente a los altos niveles de ozono porque la sequía inhibe la absorción de ozono (O_3) por parte de la vegetación (Sario et al., 2013, p. 828). El viento también es un factor importante en la dispersión,

dilución y difusión de contaminantes ya que cuando la velocidad del viento aumenta o disminuye y su dirección es irregular provoca que el resto de las partículas se mezclen y se dispersen de un área a otra (Wang, 2023, p. 6).

También es necesario tener en cuenta que evitar la expansión del COVID-19 requirió la adopción de medidas relativas a la restricción de la movilidad de los ciudadanos en un gran número de países. Numerosos estudios han analizado el impacto de las restricciones impuestas desde marzo de 2020 en el nivel de concentración de contaminantes del aire. En general, los niveles de contaminantes del aire como el dióxido de nitrógeno (NO₂), el dióxido de azufre (SO₂), el monóxido de carbono (CO) y las partículas (PM) disminuyeron globalmente en comparación con los niveles de las últimas décadas. En muchas grandes ciudades del mundo, la concentración de PM y NO₂ disminuyó en más de un 60 % durante el período de confinamiento. La proporción de reducción de los contaminantes del aire presumiblemente dependía de la rigurosidad del confinamiento. (Saha et al., 2002, pp. 5-6)

Adicionalmente es importante considerar el Efecto Fin de Semana (EFS) que se da en grandes ciudades. Durante la semana, en días laborales, se emiten las mayores cantidades de contaminantes debido al tráfico, las operaciones industriales y la movilidad urbana mientras que, los fines de semana estas emisiones se reducen a las más bajas para todos los contaminantes excepto el del Ozono (O₃). (Vasconcellos et al., 2022)

Es importante tener en cuenta los factores mencionados anteriormente para evaluar correctamente las concentraciones de contaminantes en el aire, pero especialmente, para realizar una correcta evaluación de las causas que las provocan y las medidas necesarias a aplicar para reducir las concentraciones de contaminantes en la atmósfera.

Actualmente, con la finalidad de proteger la salud y el medio ambiente, se han adoptado diversas iniciativas a nivel internacional, europeo y nacional. En cuanto a las medidas internacionales, Australia es parte del Convenio de Estocolmo sobre Contaminantes Orgánicos Persistentes desde 2004. El convenio establece una lista inicial de 12 COP, conocidos como los "Contaminantes Orgánicos Persistentes Originales" y medidas para reducir o eliminar la producción, liberación y uso de los COP, así como

para gestionar los residuos existentes de manera segura (Gobierno Australiano, 2022). A nivel nacional, se adopta el Acuerdo Nacional de Aire Limpio enfocado a coordinar la respuesta contra la contaminación de la industria y de los gobiernos tanto nacional como estatal y local (Gobierno Australiano, 2023).

Se han adoptado principalmente dos estrategias a largo plazo. Por un lado, el esquema Nacional de Informes de Energía y Efecto Invernadero (NGER) es un marco nacional único para informar de las emisiones de gases de efecto invernadero, la producción y el consumo de energía (Gobierno Australiano, 2022). Por otro lado, el mecanismo de salvaguardia requiere que los mayores emisores de gases de efecto invernadero de Australia mantengan sus emisiones netas por debajo de un límite que se reducirá progresivamente con el paso del tiempo (Gobierno Australiano, 2023).

CAPÍTULO II: METODOLOGÍA

1. MATERIALES

Para elaborar la base de datos, se han utilizado varios conjuntos de datos: el primero, relacionado con el nivel de contaminantes medido en datos horarios por las principales estaciones del servicio de meteorología del Territorio de la Capital Australiana (ACT), ubicadas en Canberra, Australia. El segundo y tercer conjunto de datos contienen indicadores relativos al clima (temperatura y precipitaciones respectivamente) recogidos por las mismas estaciones que recogen los niveles de contaminación diarios u otras estaciones muy cercanas a estas últimas. Asimismo, se ha creado otro dataset que recoge la información relativa a los incendios, y por último, un quinto conjunto de datos que recoge las variables dicotómicas o dummies relacionadas con el confinamiento por el COVID-19, y los días sábado y domingo.

Datos de la calidad del aire

En Australia, la adquisición de datos sobre la calidad del aire se lleva a cabo a través de una red de 211 estaciones de monitoreo distribuidas por todo el país. La responsabilidad del monitoreo y la publicación de datos recae sobre los gobiernos de cada

estado. En el caso de Canberra, la ciudad pertenece al Territorio de la Capital Australiana (ACT). En el ACT, la Dirección de Salud es la responsable del monitoreo de las estaciones distribuidas a lo largo del territorio. Por otro lado, el ACT ha delegado en la Agencia de Protección Medio Ambiental (EPA) la responsabilidad de realizar informes anuales de acuerdo con las medidas Nacionales de Protección Ambiental (NEPM), que proporciona los estándares de calidad del aire para los contaminantes atmosféricos más comunes. (Gobierno del ACT, 2022)

Las estaciones de calidad del aire en Australia son instalaciones remotas automáticas que registran medidas de contaminantes gaseosos y partículas. Estas estaciones deben cumplir con una serie de requisitos establecidos por las medidas NEPM que requieren que la estación de monitoreo tenga un ángulo de cielo despejado de 120° y esté situada a más de 50 metros de un camino y a 10 metros del árbol más cercano. (Emmerson, 2021)

En el ACT hay tres estaciones principales de monitoreo. Cada estación recopila datos sobre diversos contaminantes atmosféricos, como dióxido de nitrógeno (NO₂), ozono (O₃), monóxido de carbono (CO) y partículas de diferentes tamaños, como PM₁₀ y PM_{2.5}. Los sensores y dispositivos en las estaciones de monitoreo registran los niveles de contaminantes a intervalos regulares, generando datos en tiempo real o con una periodicidad específica, como por hora o por día. (Gobierno del ACT, 2023)

En este trabajo se han obtenido los datos sobre la calidad del aire durante el periodo comprendido entre el 01/01/2012 y el 14/10/2022, que, en este caso, todos ellos vienen presentados en un archivo del tipo .csv, en el que cada fila contiene información sobre la estación, la fecha y hora de la medición, y el valor correspondiente de cada partícula. Estos vienen presentados en formato diario y se muestran mediciones a distintas horas. Concretamente, se presentan los siguientes datos:

- I. **Identificación de la estación.** Se identifica cada una de las estaciones a partir de su nombre, así mismo también se incluyen las coordenadas GPS de cada estación además de la fecha y la hora a la que los datos fueron recogidos.

- II. **Contaminantes.** Se presentan los valores obtenidos de cada uno de los contaminantes atmosféricos así como el índice de calidad del aire de cada uno de ellos.

Del presente conjunto de datos, se han utilizado los datos diarios comprendidos entre 2017-2022 (se ha realizado el cálculo del valor medio diario) de las variables que se han considerado relevantes (detalladas más adelante), junto a la información relativa a la estación (Nombre y coordenadas GPS). Conviene destacar que alguna de las variables presenta un gran número de datos faltantes, motivo por el cual, serán descartadas en un futuro análisis.

Tabla 1. Estaciones de monitorización de la calidad del aire en el ACT

Nombre	Nombre Completo	Código	Año de apertura	Latitud	Longitud
Civic	Canberra Parliament House	70246	1968	-35.31	149.13
Florey	Melba (Verbrugghen Street)	70277	1973	-35.21	149.05
Monash	Tuggeranong (Isabella Plains) AWS	70339	1996	-35.42	149.09

Fuente: Elaboración propia

Datos sobre el clima

En Australia, los datos de climatología, como las precipitaciones y la temperatura, se obtienen a través de diversas fuentes y métodos. La recopilación y registro de estos datos climáticos se realiza a nivel nacional por la Oficina de Meteorología de Australia (Bureau of Meteorology). La Oficina de Meteorología opera una extensa red de estaciones meteorológicas distribuidas por todo el país. De hecho, la Bureau of

Meteorology (s.f.) explica que estas estaciones están equipadas con instrumentos y sensores especializados que registran datos climáticos, como la temperatura, la humedad, la presión atmosférica y las precipitaciones. Además, de las estaciones meteorológicas terrestres, se utilizan otros medios para recopilar datos climáticos en Australia, como por ejemplo radares, satélites o boyas marinas.

La extracción de datos tanto de la temperatura como de las precipitaciones se ha realizado mediante un criterio de proximidad a las estaciones de las que se obtiene la información relativa a los niveles de contaminación diarios. En cuanto a la temperatura, cada registro contiene la siguiente información:

- I. **Código de producto.** Se indica el código asignado por la Oficina de Meteorología para identificar y distinguir diferentes productos o conjuntos de datos meteorológicos, en este caso, la temperatura (IDCJAC0010).
- II. **Número de estación de la Oficina de Meteorología.** Se indica el código de la estación que ha recogido las temperaturas máximas diarias (por ejemplo, 70351).
- III. **Año.** Se indica el año en el que se recogió cada uno de los registros.
- IV. **Mes.** Se indica el mes en el que se recogió cada uno de los registros.
- V. **Día.** Se indica el día en el que se recogió cada uno de los registros.
- VI. **Temperatura máxima.** Se indica la temperatura máxima en grados Celsius que se registra nominalmente a las 9 am hora local. Es la temperatura más alta durante las 24 horas previas a la observación y se registra como la temperatura máxima del día anterior.

En cuanto a las precipitaciones, cada registro contiene la siguiente información:

- I. **Código de producto.** El código asignado por la Oficina de Meteorología para las precipitaciones es IDCJAC0009.
- II. **Número de estación de la Oficina de Meteorología.** Se indica el código de la estación que ha recogido el nivel de precipitación diario (por ejemplo, 70246).
- III. **Año.** Se indica el año en el que se recogió cada uno de los registros.
- IV. **Mes.** Se indica el mes en el que se recogió cada uno de los registros.

- V. **Día.** Se indica el día en el que se recogió cada uno de los registros.
- VI. **Precipitación.** Se indica el nivel de precipitación diario, las observaciones de precipitaciones diarias se realizan nominalmente a las 9 am hora local y registran el total de las últimas 24 horas. Las precipitaciones incluyen todas las formas de precipitación que llegan al suelo, como lluvia, llovizna, granizo y nieve.

En Australia, la Oficina de Meteorología cuenta con 733 estaciones operativas que midan la temperatura máxima y con 4132 estaciones operativas que midan las precipitaciones. Entre todas estas estaciones se eligieron las siguientes:

Tabla 2. Estaciones de monitorización de la temperatura máxima en el ACT

Nombre	Nombre Completo	Código	Año de apertura	Latitud	Longitud
Civic	Canberra Airport	70351	2008	-35.31	149.2
Monash	Tuggeranong (Isabella Plains) AWS	70339	1996	-35.42	149.09

Fuente: Elaboración propia

Es importante destacar que la estación Canberra Airport se trata de la estación más cercana a las estaciones Civic y Florey que recogen los niveles de contaminación. Por otro lado, la estación Tuggeranong (Isabella Plains) AWS recoge tanto los niveles de contaminación como la temperatura máxima.

Tabla 3. Estaciones de monitorización de las precipitaciones en el ACT

Nombre	Nombre completo	Código	Año de apertura	Latitud	Longitud
Civic	Canberra Parliament House	70246	1968	-35.31	149.13
Florey	Melba (Verbrugghen Street)	70277	1973	-35.21	149.05
Monash	Tuggeranong (Isabella Plains) AWS	70339	1996	-35.42	149.09

Fuente: Elaboración propia

Como se puede observar, las estaciones que recogen los niveles de precipitación diarios son las mismas que recogen los niveles de contaminación.

Datos sobre los incendios

Para obtener los datos necesarios para formar el dataset de los incendios forestales en Australia, se llevó a cabo una exhaustiva investigación en diversas páginas web oficiales. Estas fuentes proporcionan información detallada sobre los momentos en los que se han producido los incendios forestales más relevantes en el país.

Se consultaron y analizaron cuidadosamente los informes y registros proporcionados por organismos gubernamentales especializados en la gestión de incendios, como el Departamento de Medio Ambiente y Energía, y el Servicio de Bomberos y Rescate Rural de Australia. Estas instituciones cuentan con páginas web específicas donde se recopilan datos actualizados sobre los incendios forestales, incluyendo información sobre su ubicación, magnitud, duración y efectos en la calidad del aire. Además, se revisaron informes de organismos medioambientales y de protección civil de los diferentes estados y territorios de Australia, como el Country Fire Authority

(CFA) de Victoria, el Rural Fire Service (RFS) de Nueva Gales del Sur, y el Department of Fire and Emergency Services (DFES) de Australia Occidental. Estas organizaciones proporcionan datos precisos sobre los incendios forestales.

Sistema informático R para el tratamiento de los datos

R es un lenguaje de programación y una plataforma de software de código abierto al mismo tiempo. Al tener una licencia GNU GPL, su descarga y uso son gratuitos. Asimismo, cualquier persona con un mínimo de conocimientos informáticos puede acceder al código fuente de R y realizar las modificaciones que mejor se adapten a sus necesidades. (Universidad Internacional de Valencia, 2022).

Para comprender la importancia que actualmente tiene este lenguaje de programación, es necesario familiarizarse con el mismo y ver cuáles son sus principales características. Entre ellas se pueden destacar (Datademia, 2022):

- I. Diversidad de herramientas estadísticas disponibles.
- II. Compatibilidad con otros lenguajes de programación.
- III. R es un lenguaje de programación que sigue el paradigma de la programación orientada a objetos.
- IV. Capacidad de conectarse y trabajar con bases de datos.
- V. Gráficos de alta calidad.
- VI. Gran capacidad de cálculo numérico.

Una de las mayores fortalezas de R radica en su capacidad de participar en todas las etapas del análisis de datos. A continuación, se listan algunas (Universidad Internacional de Valencia, 2022):

- I. Adquisición de datos de diversas fuentes, archivos o bases de datos.
- II. Detección de valores incorrectos o duplicados para su posterior eliminación.
- III. Análisis de datos a través de la construcción de modelos de agrupamiento, clasificación y predicción.
- IV. Comunicación de los datos mediante la generación automática de informes que facilitan la exposición de los resultados y la obtención de conclusiones.

Una vez detalladas las distintas características y fortalezas de este lenguaje de programación, solo queda mencionar cuáles son algunas de las áreas potenciales para su uso (Datademia, 2022): (i) Investigación científica, (ii) Análisis estadístico, (iii) Inteligencia artificial, (iv) Machine learning, y (v) Manipulación de datos

En este proyecto, se decidió utilizar el lenguaje de programación R debido a que es uno de los lenguajes más utilizados durante el grado universitario, y porque permite no solo el análisis de los datos sino también la visualización de los mismos.

2. METODOLOGÍA PARA EL ANÁLISIS DE LA CONTAMINACIÓN

Modelo de regresión teórico

La regresión lineal múltiple es un método estadístico que se utiliza para investigar la relación entre dos conjuntos de variables: las variables dependientes e independientes. El propósito principal de la regresión múltiple es analizar y comprender cómo las variables independientes pueden explicar o predecir el comportamiento de la variable dependiente. (Carrollo, 2012, p. 2). La notación matemática de la ecuación de regresión lineal múltiple es la siguiente (Ecuación (1)):

$$Y = b_0 + b_1X_{1j} + b_2X_{2j} + \dots + b_kX_{kj} + u_j \quad (1)$$

donde “Y” es la variable dependiente, “X” son las variables independientes o explicativas, “b” son los coeficientes estimados del efecto de “X” sobre “Y”, y “u” son los residuos.

Para realizar el análisis de regresión lineal múltiple, según Rojo (2007) es preciso realizar una serie de consideraciones sobre los datos (p. 7):

- I. **Linealidad.** Existe una relación lineal entre las variables independientes y la variable dependiente puesto que los valores de esta última se generan a partir del modelo lineal: $Y = x*b + u$

- II. **Homocedasticidad.** La variabilidad de los valores de la Y para un conjunto dado de predictores es la misma independientemente de los valores de los predictores (X).
- III. **Independencia.** Las perturbaciones aleatorias son independientes entre sí.
- IV. **Normalidad.** La distribución de la perturbación aleatoria sigue una distribución normal.
- V. Obtención de las variables explicativas x_{kj} sin errores de medida.

La Universidad de Santiago de Compostela (2011) explica que los resultados que se obtienen de un modelo ajustado por regresión son los siguientes:

- I. **Estimación por mínimos cuadrados ordinarios (MCO) de los coeficientes de regresión.** Se obtienen los coeficientes “b” de la ecuación (1) que acompañan a las variables independientes y el valor de intercepción. Dichos coeficientes se obtienen a través del método de mínimos cuadrados ordinarios, que como su propio nombre indica, es aquel en el que la suma de los cuadrados entre los valores esperados y observados es mínima, en otras palabras, aquel que minimiza los residuos.
- II. **Estadísticos del ajuste.** Aquellos estadísticos que muestran lo bien que se ajusta el modelo a la realidad. Entre estos se encuentran el coeficiente de determinación R^2 y el error estándar del ajuste.
- III. **Valor del estadístico F de la regresión.** El estadístico F compara la varianza explicada por el modelo (varianza explicada por las variables independientes) con la varianza no explicada (varianza residual o error). Por tanto, sirve para testar si el modelo es globalmente significativo.
- IV. **Valor del estadístico t de la variable explicativa.** El valor del estadístico t se utiliza para evaluar la significancia individual de cada variable explicativa (coeficiente de regresión) en un modelo de regresión múltiple. El estadístico t se calcula dividiendo el coeficiente de regresión estimado para una variable independiente por su error estándar.
- V. **P-valor.** Si el p-valor asociado a una variable independiente es menor que el nivel de significancia predefinido, generalmente se considera que la variable es estadísticamente significativa y se rechaza la hipótesis nula. Esto indica que

hay evidencia suficiente para afirmar que la variable tiene un efecto significativo sobre la variable de respuesta.

Una vez realizada la regresión, es importante analizar los residuos que esta genera. Concretamente, se debe analizar que estos sean normales, independientes, y que no presenten ningún tipo de anomalía. Por lo general, para comprobar que los residuos son normales se suele realizar una prueba de normalidad, por ejemplo, el test de Kolmogorov Smirnov (Berger et al., 2014). Para comprobar que son independientes, se suele realizar a estos la prueba de Breusch-Pagan (Halunga et al., 2017).

CAPÍTULO III: APLICACIÓN DE LA METODOLOGIA

1. DESCRIPCIÓN DEL MEDIO OBJETO DE INVESTIGACIÓN

Para realizar el presente estudio, se ha seleccionado la ciudad de Canberra, en el Territorio de la Capital Australiana (ACT). La ciudad de Canberra tiene una extensión de 814,2 km² y se localiza en el sureste del país australiano a 150 km de la costa más cercana. Aunque no está directamente en la costa, se encuentra a una distancia razonable tanto del océano Pacífico como del mar de Tasmania (Gobierno del ACT, s.f.).

Canberra generalmente presenta niveles de contaminantes atmosféricos más bajos en comparación con otras grandes ciudades del mundo ya que Canberra es una ciudad planificada, lo que significa que su diseño y desarrollo fueron organizados desde un principio. El arquitecto Walter Burley Griffin fue el encargado de diseñar la ciudad en 1913, y su planificación se basó en principios de diseño urbano que buscaban integrar los espacios naturales y artificiales (Bodart, 2022). Además, Canberra tiene una población relativamente baja en comparación con otras capitales con tan solo 400.000 habitantes. Todo ello contribuye a que Canberra presente menores emisiones de contaminantes.

Las condiciones atmosféricas y geográficas de Canberra también desempeñan un papel en la calidad del aire de la región ya que la ciudad se beneficia de un clima templado con estaciones bien definidas, y una topografía montañosa con amplias áreas verdes que

contribuye a una mayor variabilidad en la calidad del aire en comparación con otras áreas urbanas.

En este sentido, el relieve de Canberra se caracteriza por estar entre la cordillera conocida como Brindabella que se extiende al oeste de la ciudad y al este de la gran cordillera divisoria. En la ciudad de Canberra, la topografía se vuelve más suave, está localizada en un valle con llanuras abiertas. En general, la altitud promedio de la ciudad de Canberra es de alrededor de 570 metros sobre el nivel del mar. Sin embargo, hay variaciones en diferentes áreas de la ciudad. Por ejemplo, en la Depresión del Molonglo, cerca del río Molonglo la altura es de 420 metros sobre el nivel del mar mientras que, al oeste, las áreas más cercanas a las montañas presentan altitudes más elevadas llegando a superar los 1.000 metros sobre el nivel del mar.

El clima de Canberra se ha calificado como un clima oceánico de relativa sequedad. En concreto, debido a su ubicación interior en el sureste de Australia, lejos de la influencia directa de los océanos, el clima se ha calificado como de montaña oceánico. Durante el invierno, que abarca los meses de junio a agosto, las temperaturas en Canberra suelen ser frescas, con mínimas que rondan los 0 °C y máximas promedio de alrededor de 12 °C. En los meses de verano (diciembre a febrero), las temperaturas diurnas promedio están alrededor de los 28 °C, aunque pueden superar los 35 °C (Bureau of Meteorology, s.f.). Además, Australia tiene algunos de los niveles más altos de radiación ultravioleta del mundo debido a la degradación de la capa de ozono del continente australiano (Cancer Council, s.f.).

En cuanto a la presencia de actividad industrial, no hay una actividad industrial significativa en la capital del país. La economía de la ciudad se basa en el sector servicios y el gobierno del país. En Canberra el principal contaminante de preocupación para la salud son las partículas que están presentes en el humo generado más comúnmente por calentadores de leña, incendios forestales, emisiones de vehículos y tormentas de polvo. Canberra se encuentra rodeada de bosques que facilitan durante la temporada anual de incendios que el humo llegue a la ciudad. Además, Canberra reúne dos condiciones que agravan la contaminación. Por un lado, Canberra es un valle rodeado de montañas lo que hace que la ciudad retenga el humo en la temporada de incendios además de promover el fenómeno conocido como inversión térmica que atrapa la contaminación durante

periodos más largos al enfriarse el aire cerca del suelo y el aire cálido situarse como una tapa sobre el que lo retiene. Por otro lado, los patrones de viento de la ciudad durante la temporada de incendios traen humo de los incendios que queda atrapado por la inversión térmica. (IQAir, 2023)

2. CONJUNTO DE DATOS

Variable dependiente: contaminantes

Las variables dependientes son aquellas cuyo valor puede estar condicionado por otras. En este caso, estas variables son aquellas relativas a los niveles de contaminación. Aunque la base de datos proporcionada por el servicio de meteorología del ACT incluye un mayor número de partículas contaminantes, con el objetivo de centrar este estudio, la evaluación de la calidad del aire se ha realizado para los índices de contaminación NO_2 , O_3 y PM_{10} , durante el periodo comprendido entre el 01/01/2017 y 14/10/2022.

- **Dióxido de nitrógeno (NO_2).** Es un gas tóxico que se genera principalmente por la combustión de combustibles fósiles, por ejemplo, en los motores de vehículos, centrales eléctricas y procesos industriales (Kumar, 2010). De acuerdo con la EPA (2023), la exposición a altos niveles de NO_2 puede causar problemas respiratorios, agravar enfermedades cardiovasculares, afectar a la mucosa de los ojos, nariz, garganta, etc.
- **Ozono (O_3).** A nivel del suelo, el ozono es un contaminante y un componente del smog. Se forma por la reacción química de los óxidos de nitrógeno (NO_x) y los compuestos orgánicos volátiles (COV) en presencia de la luz solar (Departamento de Salud y Control Ambiental de Carolina del Sur, s.f.). De acuerdo con la EPA (2023) la exposición al O_3 puede irritar y dañar los tejidos del sistema respiratorio, así como contribuir al desarrollo de enfermedades cardiovasculares y aumentar la mortalidad en casos de exposición prolongada.
- **PM_{10} .** Se tratan de partículas suspendidas en el aire con un diámetro aerodinámico menor o igual a 10 micrómetros. Estas partículas son emitidas por diversas

fuentes, como la quema de combustibles fósiles, incendios forestales, la actividad industrial, los vehículos motorizados, las actividades de construcción y el polvo del suelo. (Gobierno del Estado de California, s.f.). Según la EPA (2022), numerosos estudios científicos han relacionado la exposición a este tipo de partículas con diversos problemas de salud, por ejemplo, muerte prematura en personas con enfermedades cardíacas o pulmonares, infartos no mortales, ritmo cardíaco irregular, etc.

Variables independientes

Las variables independientes seleccionadas han sido la temperatura máxima diaria, la precipitación media, la existencia de incendios, cuatro periodos correspondientes al COVID-19 con relevancia significativa, y si el día es sábado o domingo.

Las variables independientes utilizadas en los modelos de regresión son:

Tabla 4. Variables independientes utilizadas en los modelos de regresión

NOMBRE	TIPO	DESCRIPCIÓN	UNIDADES
Temperatura	Numérica continua	Temperatura máxima diaria	Grados Celsius
Precipitación media	Numérica continua	Precipitación diaria media de 07:00 horas a 07:00 horas	Milímetros (mm)
Incendios	Dicotómica	Los incendios de mayor impacto ocurridos en Australia	1 / 0
Fase 0	Dicotómica	El gobierno del ACT declara una emergencia de salud pública, desde el 16/03/2020 hasta el 01/05/2020	1 / 0
Fase 1	Dicotómica	Periodo de desescalada desde el 02/05/2020 hasta el 30/06/2020	1 / 0
Confinamiento	Dicotómica	El gobierno del ACT decretó un confinamiento desde el 12/08/2021 hasta el 15/10/2021	1 / 0
Desconfinamiento	Dicotómica	Periodo de desescalada desde el 16/10/2021 hasta el 16/12/2021	1 / 0
Sábado	Dicotómica	Sábado	1 / 0
Domingo	Dicotómica	Domingo	1 / 0

Fuente: Elaboración propia

Variables climatológicas. Los factores meteorológicos como la temperatura ambiente, la humedad, las precipitaciones, velocidad y dirección del viento entre otros pueden influir en los contaminantes atmosféricos tras la emisión de estos últimos. Los contaminantes tras ser vertidos por una fuente móvil o estacionaria pasan por distintos procesos de transformación química, deposición, transporte y dispersión en la atmósfera, siendo estos dos últimos procesos a los que afectan las variables climatológicas mencionadas anteriormente. (Alahmad et al., 2023). Por consiguiente, se han considerado la temperatura máxima diaria y el nivel de precipitación medio diario como variables relevantes a analizar.

Respecto a la temperatura, Alahmad et al., (2023) explican que las altas temperaturas favorecen la reacción química y la formación de contaminantes secundarios, como el ozono troposférico (O_3). Asimismo, afirman que las altas temperaturas incrementan el potencial oxidante de los componentes atmosféricos para producir mayor cantidad de partículas de sulfato, las cuales son uno de los principales constituyentes de las partículas finas en suspensión (PM). Por otro lado, las bajas temperaturas pueden favorecer la acumulación de contaminantes en niveles cercanos al suelo debido al fenómeno conocido como inversión térmica. Durante este proceso, el aire frío se acumula cerca del suelo, mientras que el aire más cálido actúa como una especie de tapa que atrapa los contaminantes en la capa inferior de la atmósfera por períodos más prolongados (Guadalupe et al., 2012).

En cuanto a las precipitaciones, es uno de los principales procesos naturales que puede contribuir a mejorar la calidad del aire a través de un efecto de lavado sobre los contaminantes atmosféricos situados por debajo de las nubes, a los que arrastra hacia el suelo. (Duhanyan and Roustan, 2011)

Variables relacionadas con el efecto del fin de semana y las fuentes naturales de contaminación. Los días del fin de semana, es decir, los sábados y domingos, muestran patrones distintos en los contaminantes atmosféricos en comparación con su comportamiento de lunes a viernes, especialmente en zonas urbanas (Sicard et al., 2020). El hecho que durante los fines de semana se reduzca el tráfico vehicular y la actividad industrial puede provocar la reducción de las emisiones de gases contaminantes, como por ejemplo, de los óxidos de nitrógeno (NO_x) y de los compuestos orgánicos volátiles

(COV). Sin embargo, existe el llamado “efecto ozono de fin de semana” ya que como consecuencia de la reducción de los contaminantes NO_x y de los COV, las concentraciones de O_3 aumentan los fines de semana. (Seguel, 2012; Sicard et al., 2020)

En relación a las fuentes naturales de contaminación, los incendios forestales son un fenómeno natural, pero en las últimas décadas, la frecuencia e intensidad de los incendios han aumentado debido al aumento de las temperaturas a consecuencia del cambio climático y se prevé que en 2050 los incendios aumenten un 30%. (Naciones Unidas [UN], 2022)

El aumento de la frecuencia e intensidad con la que se producen los incendios suponen riesgos para los ecosistemas y para la salud humana. Como explican Graham et al. (2021), los incendios liberan una amplia gama de contaminantes en la atmósfera, entre ellos, cantidades de dióxido de carbono (CO_2), monóxido de carbono (CO), óxidos de nitrógeno (NO_x) y compuestos orgánicos volátiles (COV). Estos gases contribuyen al calentamiento global, el deterioro de la calidad del aire y la formación de smog. Los incendios también emiten partículas finas ($\text{PM}_{2,5}$) y (PM_{10}) que pueden transportarse largas distancias. Estas partículas son perjudiciales para la salud humana, ya que pueden penetrar en los pulmones y causar problemas respiratorios, irritación de los ojos y agravar enfermedades cardiovasculares. Un estudio de la Junta de Recursos del Aire de California (2021) encontró que, en los incendios de California de 2018, además de liberarse los contaminantes previamente mencionados, se liberaron altas cantidades de plomo y hierro (p. 7).

En concreto, en el caso de Australia, este país es conocido por su propensión a los incendios forestales debido a sus condiciones climáticas y geográficas. Australia experimenta un clima propenso a los incendios forestales, con veranos calurosos y secos que aumentan la disponibilidad de material inflamable. Los patrones de sequía prolongada y los eventos climáticos extremos, como El Niño, agravan aún más la situación, creando condiciones ideales para la propagación de incendios de gran magnitud (Rahman et al., 2018). A parte de la pérdida de hábitats, el humo emitido por los incendios inunda las principales ciudades del país, lo que influye negativamente en la calidad del aire y en la salud de sus habitantes (Graham et al., 2021).

Canberra se encuentra entre estas ciudades expuestas al humo de los incendios ya que se encuentra rodeada por áreas boscosas y parques nacionales, lo que aumenta el riesgo de incendios forestales en las cercanías de la ciudad (Williamson et al., 2022). Además, la capital de Australia, a pesar de encontrarse entre las ciudades con mejor calidad del aire, ha ocupado durante los incendios de diciembre de 2019- enero de 2020 un puesto entre las ciudades con peor calidad del aire del mundo (Vassarotti, 2021, p. 18).

VARIABLES RELACIONADAS CON EL COVID-19. El COVID-19 ha traído consigo restricciones en la movilidad como el confinamiento o el cierre de negocios, y por consiguiente un descenso de los factores que emiten contaminantes a la atmósfera. En Australia, la respuesta ante el COVID-19 fue distinta en los diferentes estados que conforman el país. El objeto de este trabajo se limita al estudio de Canberra, por lo que las restricciones que se analizan son aquellas impuestas por el ACT.

El primer periodo seleccionado (denotado como “Fase 0”) abarca desde el 16 de marzo de 2020 cuando el gobierno del ACT declara la emergencia para la salud pública hasta el 1 de mayo de 2020 cuando el gobierno comienza a aliviar todas las restricciones impuestas. Durante este periodo se cerraron todos aquellos negocios y actividades no esenciales, se cerraron las escuelas, y se limitó la asistencia a hospitales a casos de urgencia. (Storen y Corrigan, 2020)

El segundo periodo seleccionado (denotado como “Fase 1”) abarca desde el 2 de mayo de 2020 cuando el gobierno del ACT comenzó a retirar las restricciones impuestas hasta el 30 de junio de 2020 cuando se eliminó la última restricción relativa al número de personas que podían reunirse en espacios cerrados. Durante este periodo se retiraron todas las restricciones que habían sido impuestas a raíz del COVID-19: se abrieron de nuevo los negocios no esenciales y se decretó el regreso a la escuela de manera presencial. (Storen y Corrigan, 2020)

El tercer periodo seleccionado (denotado como “Confinamiento”) corresponde al primer y único confinamiento decretado en el ACT. El confinamiento del ACT se inició el 12 de agosto de 2021 y finalizó el 15 de octubre de ese mismo año. Durante el confinamiento, los residentes de Canberra solo podían abandonar sus casas para la compra de alimentos, recibir atención médica esencial, realizar ejercicio durante 1 hora

por día, recibir cuidados esenciales y para realizar trabajos esenciales. (Edwards et al., 2022)

El cuarto periodo escogido (denotado como “Desconfinamiento”) abarca desde el 16 de octubre de 2021 cuando se decretó el fin del confinamiento hasta el 12 de diciembre de 2021. El gobierno de Australia diseñó un plan por pasos para la vuelta a la normalidad tras el confinamiento de un estado. Cada estado tenía la potestad de decidir cuándo se implementaría cada paso. En el caso del ACT, la eliminación de prácticamente todas las restricciones, incluida la obligatoriedad del uso de mascarilla, se produjo el 12 de diciembre de 2021. (Edwards et al., 2022)

3. TRATAMIENTO DE DATOS CON R

Con el objetivo de obtener la base de datos definitiva con las variables mencionadas anteriormente, se realizó en primer lugar una limpieza y filtrado de los datos de contaminación utilizando *RStudio*. Además, para el análisis de regresiones se seleccionó el periodo comprendido entre el año 2017 y 2022, y se consideraron los contaminantes NO₂, PM₁₀ y O₃.

En primer lugar, puesto que el dataset tiene registros diarios y por horas, se calculó el valor medio diario de la concentración de cada una de las partículas mencionadas anteriormente para cada una de las estaciones. Asimismo, los datos faltantes (denotados como “NA”) se estimaron utilizando el método de regresiones de la librería *mice* de R. Finalmente, los datos procesados se agruparon por estación y partícula en estructuras de datos tipo dataframe mediante la función creada "estacion".

En segundo lugar, se obtuvieron los datos diarios relacionados con las condiciones climatológicas, tales como la temperatura y las precipitaciones, obtenidos de las mismas estaciones que registran los niveles de contaminación o de las estaciones meteorológicas más cercanas a las 3 principales estaciones de monitoreo de calidad del aire en Canberra. Estos datos estaban en archivos de formato .csv, por lo que se convirtieron a formato Excel.

En tercer lugar, se elaboró una cuarta base de datos en Excel con el propósito de recopilar información acerca de los incendios forestales más grandes y de mayor impacto ocurridos en Australia. Para completar el proceso, se elaboró una base de datos relativa al confinamiento y a las restricciones de movilidad por el COVID-19 en función de las noticias publicadas en páginas oficiales del gobierno australiano.

Por último, una vez preparados los datos, se fusionaron los conjuntos de datos utilizando la variable Fecha como referencia, ya que es común en los cinco conjuntos (Calidad del aire, Temperatura, Precipitaciones, Incendios y COVID-19), generando conjuntos de datos separados para cada estación: Monash, Civic y Florey.

Como resultado de todo este proceso, se obtuvieron un total de 8 archivos Excel, uno por cada estación y contaminante, que contendrán 2113 registros correspondientes a los datos diarios entre 2017 y 2022. Es importante destacar que el total de archivos Excel no es 9 porque en la estación Civic no se disponen de los registros relativos al NO₂, por tanto, no se ha podido estudiar los niveles de contaminación de esta partícula en esta estación. Además, cada uno de los archivos Excel mencionados anteriormente incluía las siguientes variables: (i) Fecha, (ii) Nivel medio diario del contaminante, (iii) Temperatura máxima, (iv) Precipitación media, (v) Incendios, (vi) Fase 0, (vii) Fase 1, (viii) Confinamiento, (ix) Desconfinamiento, (x) Sábado, y (xi) Domingo.

CAPÍTULO IV: RESULTADOS

1. CONSTRUCCIÓN Y RESULTADOS DE LOS MODELOS ECONÓMICOS DE REGRESIÓN

En el presente capítulo se presentan los resultados del análisis descriptivo univariado, análisis de correlación y análisis de regresión lineal múltiple de las partículas contaminantes NO₂, O₃ y PM₁₀ y las variables independientes.

La concentración del contaminante c en la estación se modeló como una función de las variables independientes (Ecuación (2)):

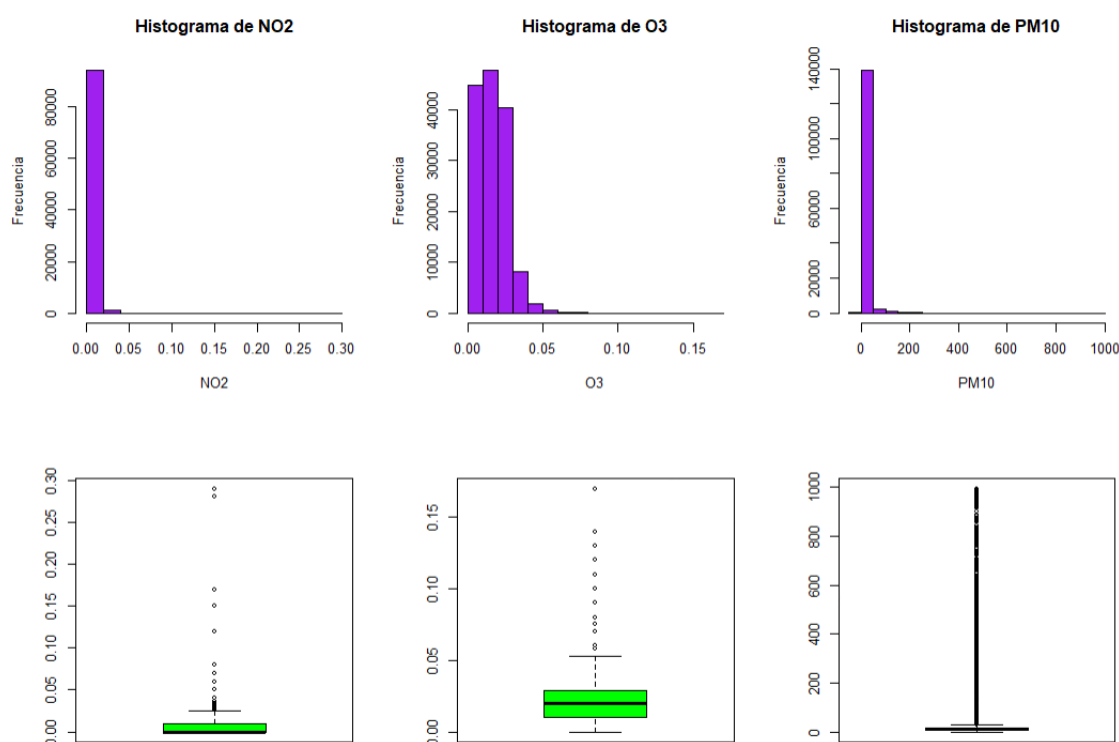
$$\ln(P_{sct}) = \alpha_{sc} + \beta_{sc}D + \gamma_{sc}F + \varepsilon_{sc} \quad (2)$$

donde P_{sct} es el nivel medio diario del contaminante c en la estación s y en el día t , D es un vector de variables dicotómicas que toman valores 0 o 1, F es un vector que contiene las variables de control continuas relativas a valores climatológicos y ε_{sc} es un término de error.

El parámetro β_{sc} o impacto marginal recoge en términos porcentuales el efecto de la presencia de cada una de las variables dicotómicas en el contaminante c medido en la estación s , del vector D : Incendio, Fase 0, Fase 1, Confinamiento, Desconfinamiento, Sábado, Domingo. Por su parte, el intercepto y el efecto de las variables continuas restantes es capturado por α y γ respectivamente. En este sentido, γ_{sc} recoge el efecto porcentual en la concentración del contaminante c medido en la estación s que tiene el incremento en una unidad de cada una de las variables continuas del vector F : Temperatura máxima y Precipitación media.

Para cada uno de los distintos contaminantes (NO_2 , O_3 y PM_{10}) se estimó una ecuación separada. Dichas ecuaciones se ajustaron a los datos mediante el método MCO. Además, se decidió tomar logaritmos de los niveles de contaminación ya que su distribución era sesgada asimétrica positiva, con cola a la derecha, como se puede observar en la Figura 1, y había valores atípicos, especialmente en la variable PM_{10} , que posteriormente se decide mantener, ya que esos altos valores de concentración de PM_{10} podrían ser consecuencia de un incendio de gran magnitud que ocurrió en enero del 2020 en Australia. Adicionalmente, al tomar logaritmos, los efectos estimados podrían interpretarse en términos porcentuales.

Figura 1. Gráfico de la distribución de los datos por cada partícula



Con el objetivo de identificar los patrones generales presentes en los datos, se realizó un análisis descriptivo de cada partícula con cada una de las variables independientes. Por un lado, para las variables dicotómicas que son de tipo factor, se usaron “diagramas de cajas” y, por otro lado, para las variables continuas se calcularon las correlaciones. Los diagramas de cajas de la concentración de las partículas con cada variable factor alertan de la presencia de valores atípicos que superan la varianza esperada y que podrían afectar los resultados del modelo de regresión.

1.1 Resultados de los modelos de regresión en Monash

Partícula NO₂

Como análisis descriptivo previo, para ver el comportamiento de los datos, la Figura 2 muestra la posible existencia de valores atípicos de la concentración promedio de NO₂ por cada variable independiente de tipo factor, así mismo, en la Figura 3 se observa que el promedio de concentración de NO₂ tiene una correlación negativa y significativa con la temperatura ($r = -0,24$) y las precipitaciones ($r = -0,16$), por tanto, a

medida que las concentraciones de NO_2 aumentan, la temperatura baja y las precipitaciones son menores, por otro lado, en el diagrama de dispersión, se observan claramente unos puntos atípicos que será conveniente excluir para poder ajustar el modelo más eficientemente.

Figura 2. Diagrama de cajas de la concentración promedio de NO_2 versus las variables independientes de tipo factor en Monash

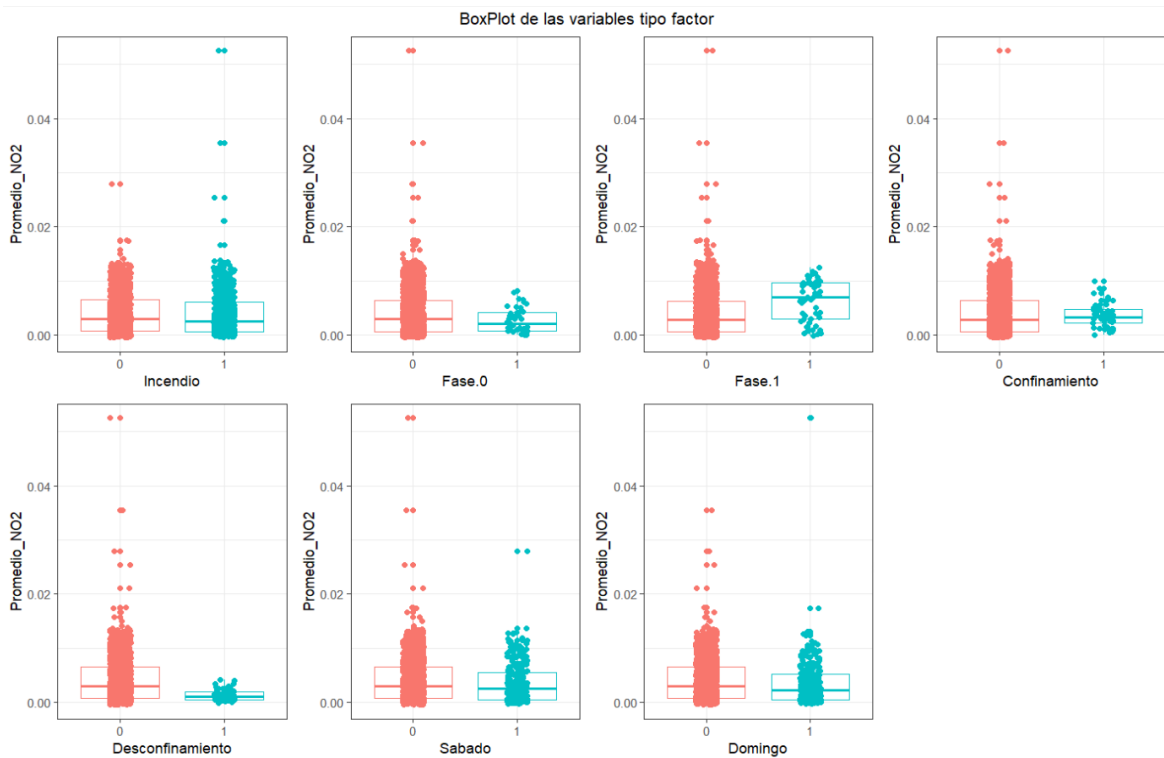
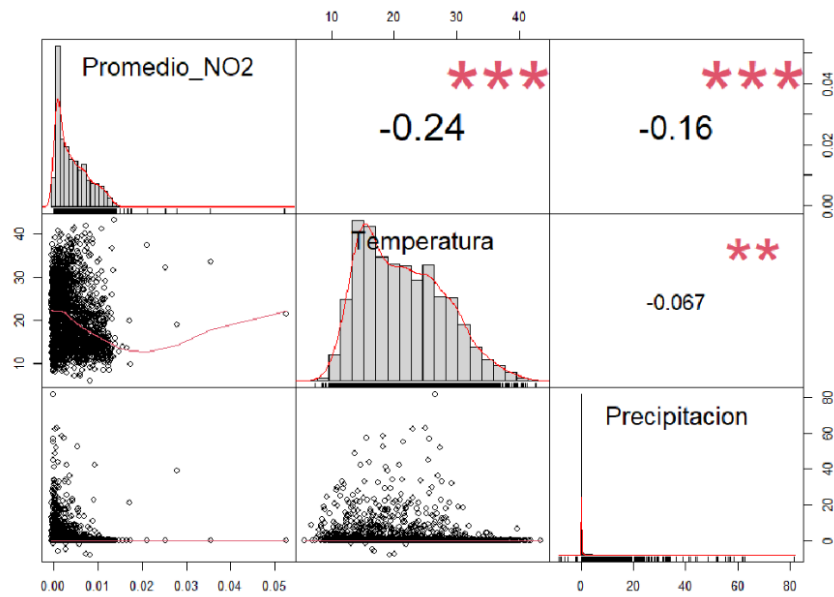
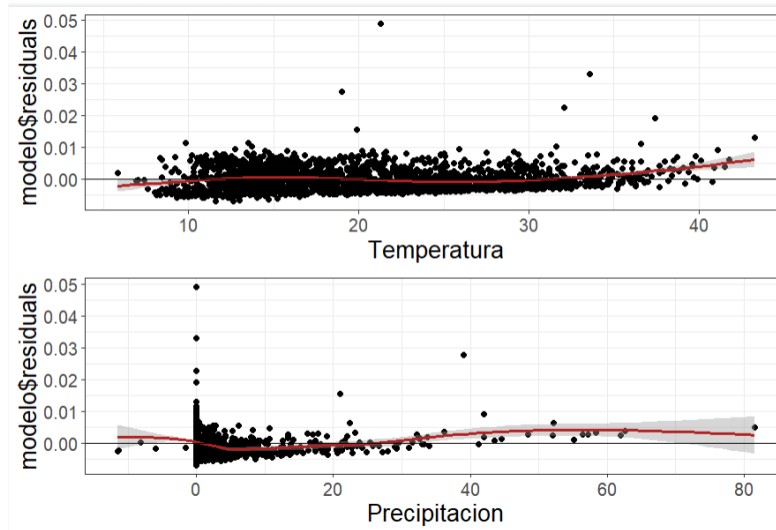


Figura 3. Correlación de la concentración promedio de NO₂ versus la temperatura y la precipitación en Monash



Para encontrar el mejor modelo de regresión posible, en un principio se ajustó un primer modelo general, en el que se obtiene que la temperatura, las precipitaciones, la fase 0, la fase 1, el desconfiamiento, y los fines de semana, son variables significativas ($p\text{-valor} < 0,05$) para explicar la variabilidad de la concentración promedio de NO₂, no obstante, como análisis adicional se evaluaron los supuestos de linealidad, normalidad y homocedasticidad. En la Figura 4 se puede observar que los residuos se distribuyen aleatoriamente en torno a 0, por lo que se puede afirmar que se cumple el supuesto de linealidad.

Figura 4. Comprobación del supuesto de linealidad del modelo de regresión de NO_2 en Monash



Respecto a la normalidad, en la Figura 5 se observa que la mayoría de los puntos se distribuyen sobre la diagonal, aunque se observan algunos valores atípicos, siendo el test de Kolmogorov Smirnov el que confirma que no se cumple este supuesto ($p\text{-valor}= 1,697e-09 < 0,05$). Por último, respecto a la homocedasticidad, el gráfico de variabilidad constante (Figura 6) de los residuos muestra que la dispersión de los puntos tiene un patrón similar a una cónica, no obstante, hay otros puntos que están dispersos, como consecuencia, los residuos no son dependientes del valor ajustado y por lo tanto se cumple el supuesto de homocedasticidad, así mismo el estadístico de Breusch-Pagan lo confirma ($p\text{-valor}= 0,4347 > 0,05$).

Figura 5. Comprobación del supuesto de normalidad del modelo de regresión de NO_2 en Monash

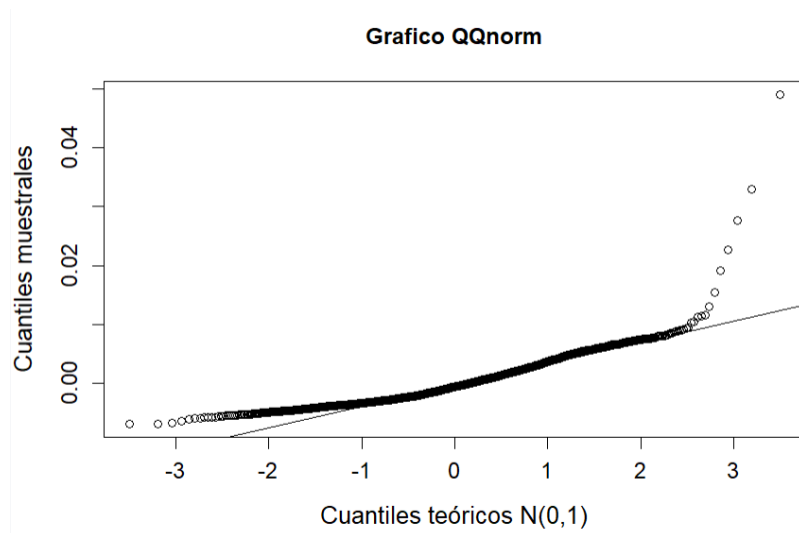
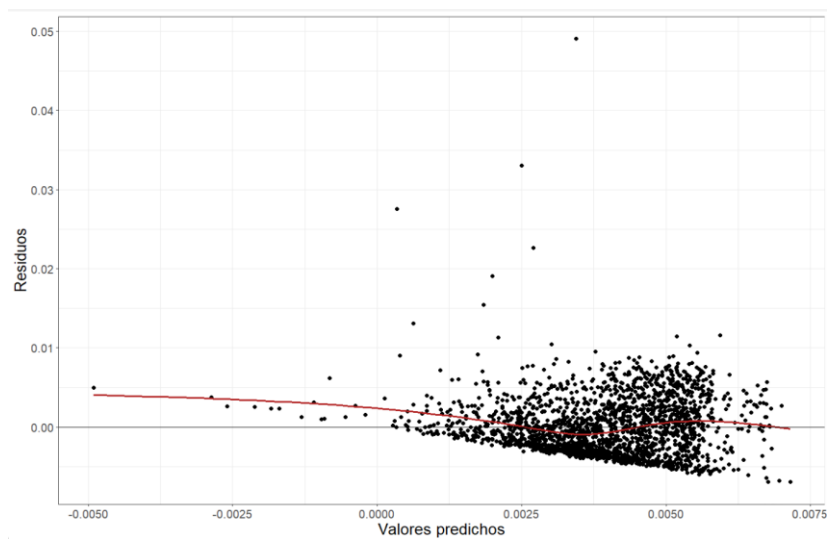
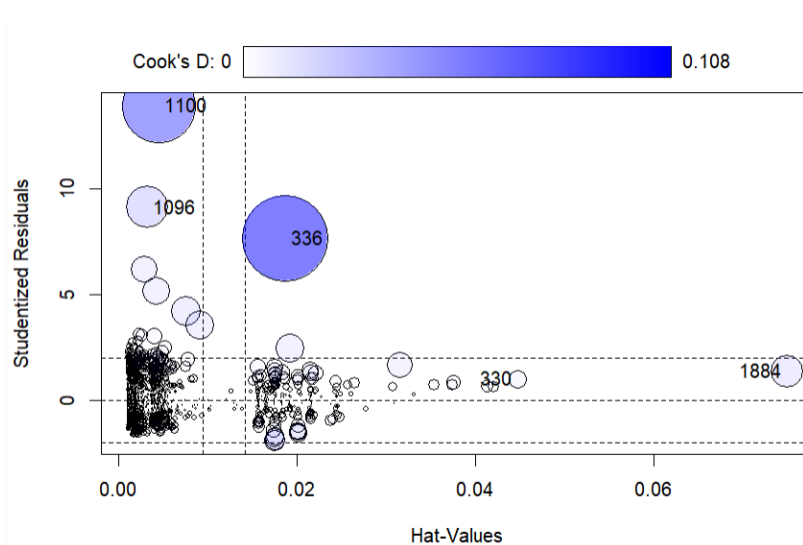


Figura 6. Comprobación del supuesto de homocedasticidad del modelo de regresión de NO_2 en Monash



Dado que se incumple el supuesto de normalidad, se procedió a detectar los valores atípicos con la función “outlierTest” y con la función “influencePlot” los valores influyentes (Figura 7), que utiliza el estadístico de la distancia de Cook para medir la diferencia entre el modelo completo con todas las muestras menos el modelo generado eliminando la muestra bajo estudio.

Figura 7. Valores influyentes en el modelo de concentración promedio de NO_2



Una vez excluidos de la base de datos los puntos atípicos e influyentes, se ajustó de nuevo un modelo de regresión con el método “stepwise mixto” para elegir el mejor modelo en función del indicador de Akaike (AIC). Respecto al modelo final, el error estándar medio del modelo indica que los resultados de la muestra son representativos de la población dado que se sitúan por debajo de 0,5 (Residual standard error: 0,2286). En primer lugar, respecto a las variables significativas y que influyen en el coeficiente de Akaike, los coeficientes indican que por cada grado centígrado que aumente la temperatura, disminuyen las concentraciones de NO_2 en un 0,96% y por cada milímetro de lluvia, disminuyen las concentraciones de NO_2 en un 0,82%. Asimismo, durante los incendios y la fase 0, el promedio de concentración de NO_2 disminuyó en un 2,77% y en un 8,54% respectivamente. Asimismo, durante el desconfinamiento disminuyó en un 17,24%. Finalmente, los sábados y domingos, se observó una reducción del 4,25% y del 5,62% respectivamente en la concentración de NO_2 .

Por último, el confinamiento no es una variable significativa (p -valor > 0,05), pero sí que influye en el coeficiente de Akaike. Se ha obtenido que durante el confinamiento la concentración de NO_2 disminuye en un 4,33%.

Tabla 5. Resultados del modelo de regresión para la partícula NO₂ en Monash

Variabes	Coefficientes	Error std.	Estadístico t	P-valor
INTERCEPTO	-4,0689	0,0167	-242,991	<2e-16
TEMPERATURA	-0,0096	0.0007	-13,201	<2e-16
PRECIPITACIÓN	-0,0082	0.0008	-9,944	<2e-16
INCENDIO	-0,0278	0.0110	-2,516	0.0119
FASE 0	-0,0854	0.0340	-2,516	0.0119
FASE 1	0,1018	0.0304	3,346	0.0008
CONFINAMIENTO	-0,0433	0,0292	-1,482	0.1385
DESCONFINAMIENTO	-0,1724	0,0297	-5,800	7.62e-09
SÁBADO	-0,0425	0,0145	-2,937	0.0033
DOMINGO	-0,0562	0,0145	-3,881	0.0001

Fuente: *Elaboración propia*

Partícula O₃

En relación a la concentración de O₃ en Monash, la Figura 8 permite observar la posible presencia de valores atípicos para cada variable predictora de tipo factor, así mismo en la Figura 9 se observa que el promedio de concentración de O₃ tiene una correlación positiva y significativa con la temperatura ($r= 0,39$) y las precipitaciones ($r= 0,073$). Esto implica que a medida que las concentraciones de O₃ aumentan, la temperatura y las precipitaciones aumentan, por otro lado, en el diagrama de dispersión, se observa claramente unos puntos influyentes y atípicos que será conveniente eliminar para lograr un mejor ajuste del modelo.

Figura 8. Diagrama de cajas de la concentración promedio de O₃ versus las variables independientes de tipo factor en Monash

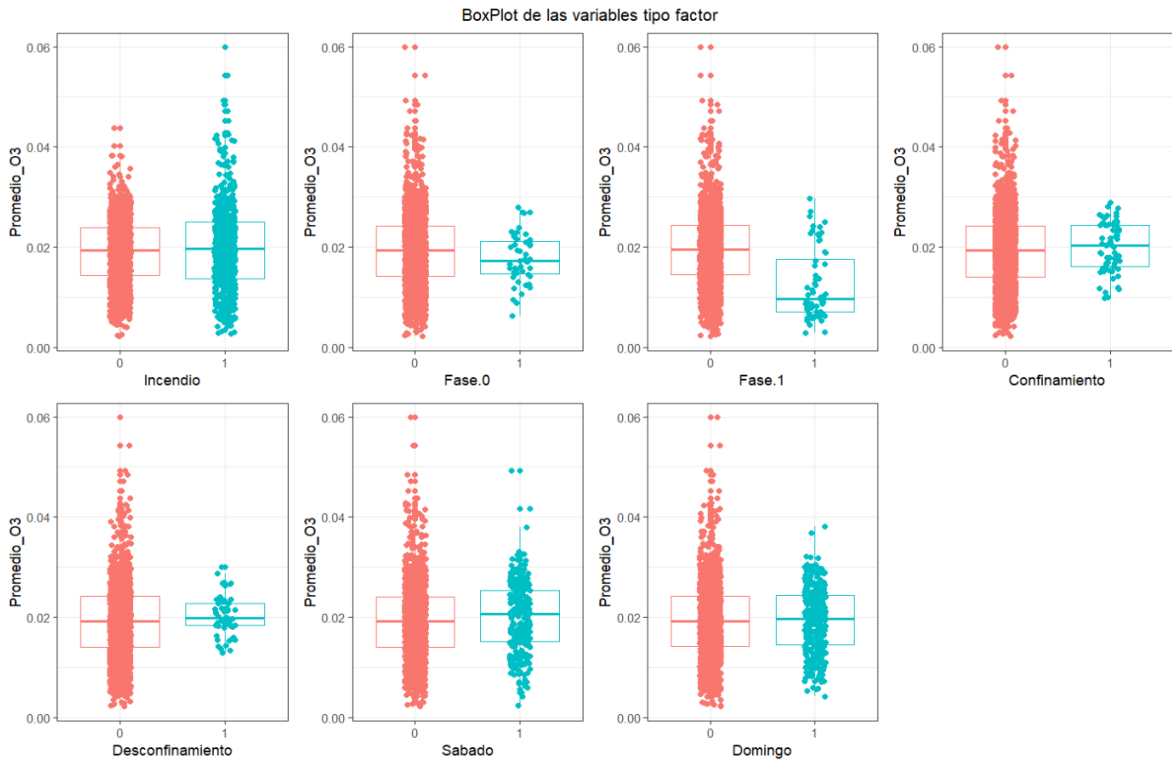
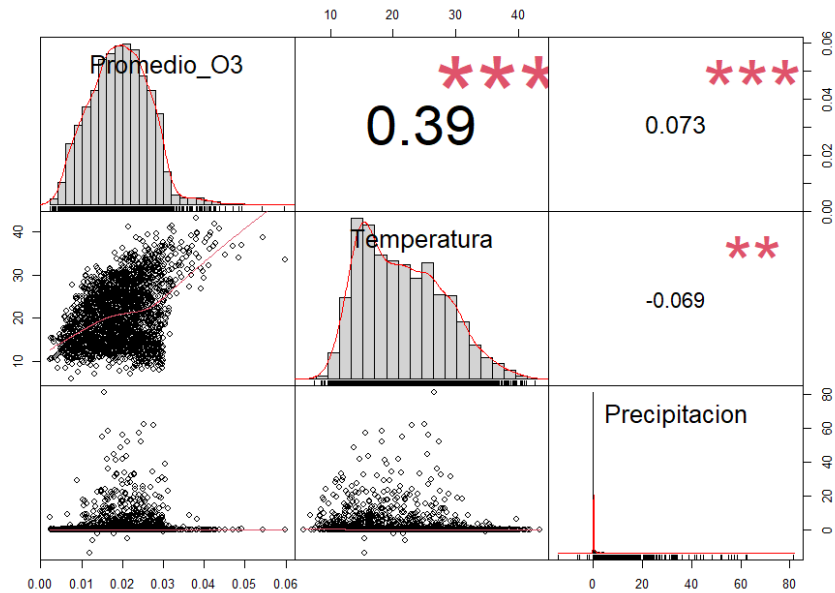


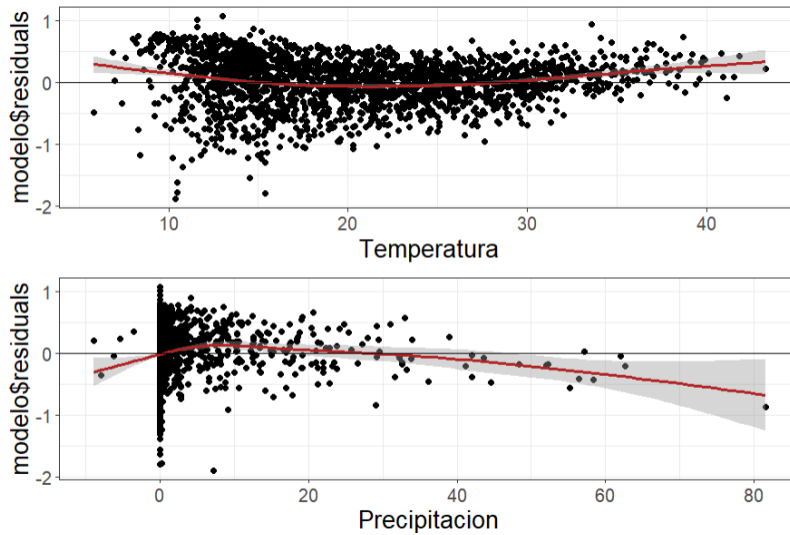
Figura 9. Correlación de la concentración promedio de O₃ versus la temperatura y la precipitación en Monash



En el primer modelo general ajustado se obtuvo que la temperatura, las precipitaciones, la fase 1, el confinamiento y los sábados, son variables significativas que ayudan a explicar la concentración de O₃ en Monash ($p\text{-valor} < 0,05$). En cuanto a los

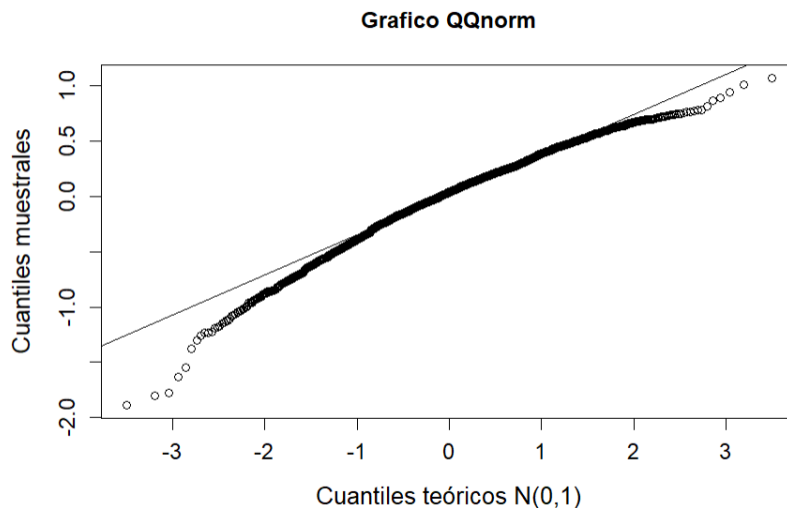
supuestos de linealidad, normalidad y homocedasticidad, como se observa en la Figura 10, los residuos se distribuyen aleatoriamente en torno a 0, lo que indica que se cumple el supuesto de linealidad.

Figura 10. Comprobación del supuesto de linealidad del modelo de regresión de O_3 en Monash



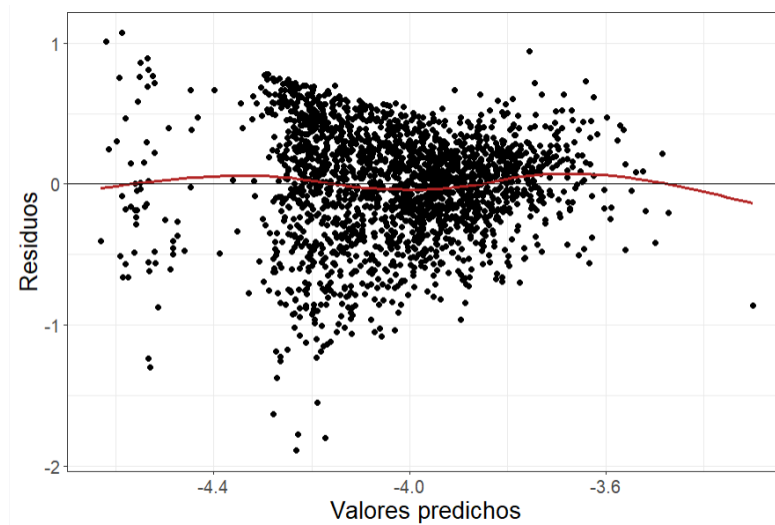
Sin embargo, respecto a la normalidad, en la Figura 11 se observa que la mayoría de los puntos se distribuyen sobre la diagonal (se observan algunos valores atípicos), siendo el test de Kolmogorov Smirnov el que confirma que no se cumple este supuesto ($p\text{-valor} = 0,0003081 < 0,05$).

Figura 11. Comprobación del supuesto de normalidad del modelo de regresión de O_3 en Monash



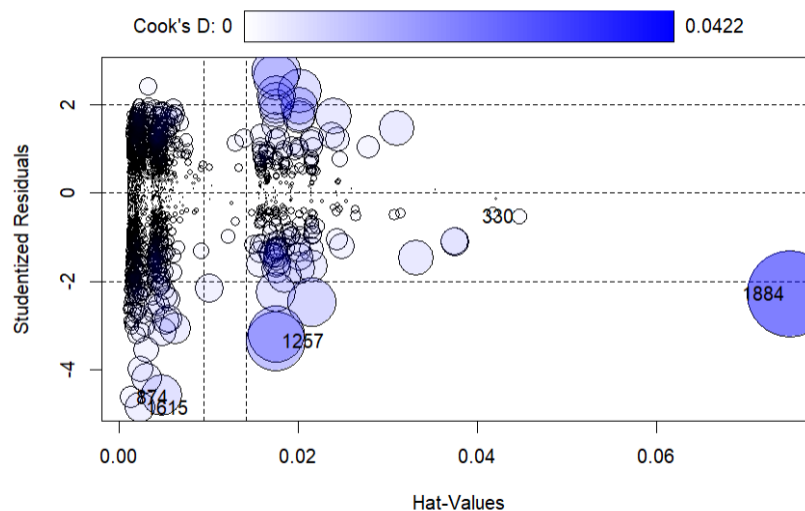
Respecto a la homocedasticidad, el gráfico de variabilidad constante de los residuos (Figura 12) muestra que la dispersión de los puntos tiene un patrón parecido a una cónica, lo que indica que la variabilidad es dependiente del valor ajustado y, por lo tanto, no hay homocedasticidad. Asimismo, el estadístico de Breusch-Pagan también lo confirma (p-valor= $2.2e-16 < 0,05$).

Figura 12. Comprobación del supuesto de homocedasticidad del modelo de regresión de O_3 en Monash



Dado que se incumplen los supuestos de normalidad y homocedasticidad, se procedió a detectar los valores atípicos con la función “outlierTest” y con la función “influencePlot” los valores influyentes (Figura 13).

Figura 13. Valores influyentes en el modelo de concentración promedio de O_3



Una vez excluidos de la base de datos los puntos atípicos e influyentes, se ajustó nuevamente un modelo de regresión en el método “stepwise mixto” para elegir el mejor modelo en función del indicador de Akaike (AIC). En este nuevo modelo, se obtuvo el error estándar medio por debajo de 0,5 (Residual standard error: 0,3848), por tanto, los resultados de la muestra son representativos de la población. En primer lugar, respecto a las variables significativas y que afectan al coeficiente de Akaike, los coeficientes indican que la temperatura y las precipitaciones influyen en el incremento de concentración de O₃, ya que por cada grado centígrado que aumente la temperatura, las concentraciones de O₃ aumentan en un 2,22% y por cada milímetro de lluvia las concentraciones de O₃ aumentan en un 0,82%. Durante la fase 1, se observó una disminución promedio del 34,28% mientras que tanto en el confinamiento se observó un aumento promedio del 14,98%. Finalmente, los sábados se observó un aumento promedio de aproximadamente el 5,11%.

En segundo lugar, es importante destacar que el desconfinamiento no es una variable significativa (p-valor > 0,05), pero influye en el coeficiente de Akaike, en este periodo se registró un aumento del 9,44%. Por último, los incendios, la fase 0 y los domingos ni son variables significativas ni afectan al coeficiente de Akaike.

Tabla 6. Resultados del modelo de regresión para la partícula O₃ en Monash

Variables	Coefficientes	Error std.	Estadístico t	P-valor
INTERCEPTO	-4,5068	0,0275	-164,006	< 2e-16
TEMPERATURA	0,0223	0,0012	18,328	< 2e-16
PRECIPITACIÓN	0,0082	0,0013	6,097	1.29e-09
FASE 1	-0,3428	0,0515	-6,656	3.57e-11
CONFINAMIENTO	0,1498	0,0488	3,070	0.0022
DESCONFINAMIENTO	0,0944	0,0497	1,899	0.0576
SABADO	0,0511	0,0240	2,128	0.0335

Fuente: Elaboración propia

Partícula PM_{10}

Respecto a la concentración de partículas de PM_{10} en Monash, la Figura 14 indica la posible existencia de valores atípicos para cada variable predictora de tipo factor, así mismo en la Figura 15 se observa que el promedio de concentración de PM_{10} tiene una correlación positiva y significativa con la temperatura ($r= 0,20$) y una correlación significativa y negativa con las precipitaciones ($r= -0,078$). Esto sugiere que a medida que las concentraciones de PM_{10} aumentan, la temperatura aumenta y las lluvias disminuyen, por otro lado, en el diagrama de dispersión, se observan claramente puntos influyentes y valores atípicos que será conveniente excluir para lograr un mejor ajuste del modelo.

Figura 14. Diagrama de cajas de la concentración promedio de PM_{10} versus las variables independientes de tipo factor en Monash

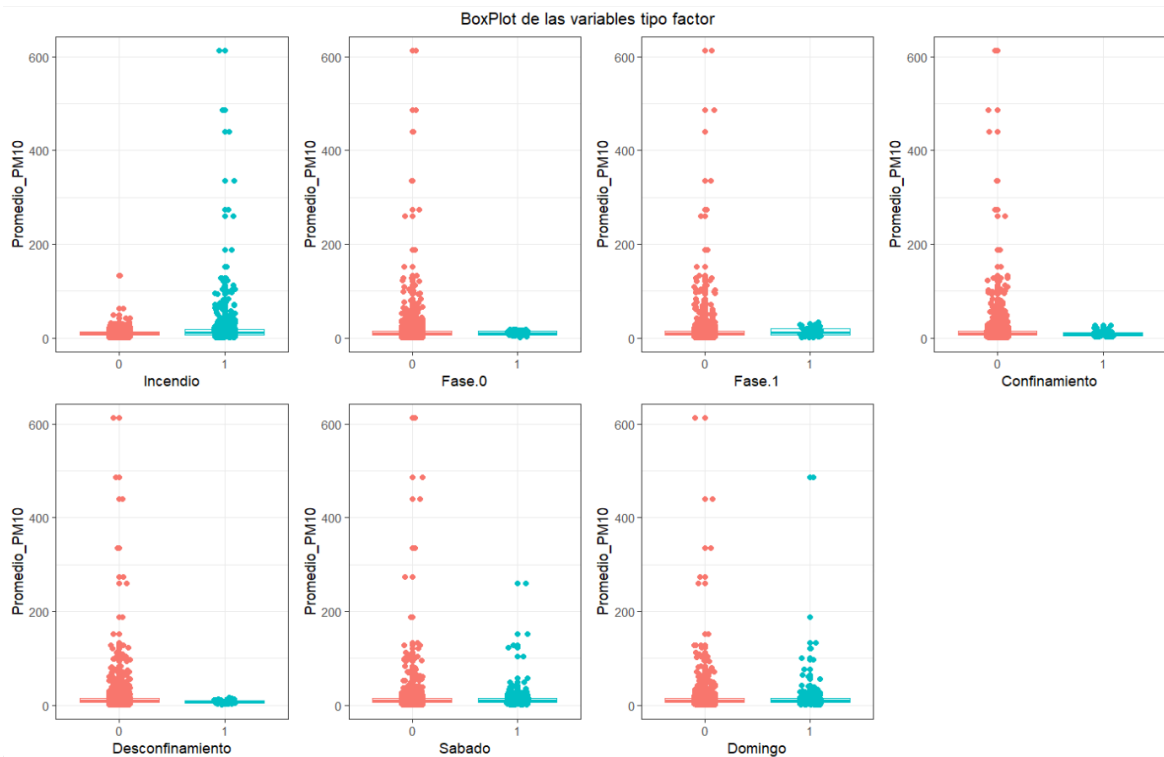
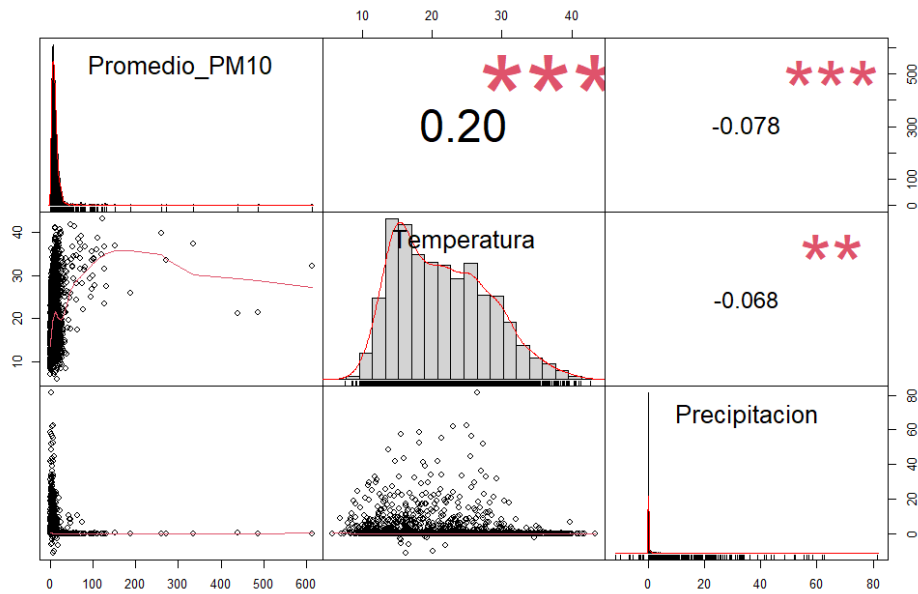
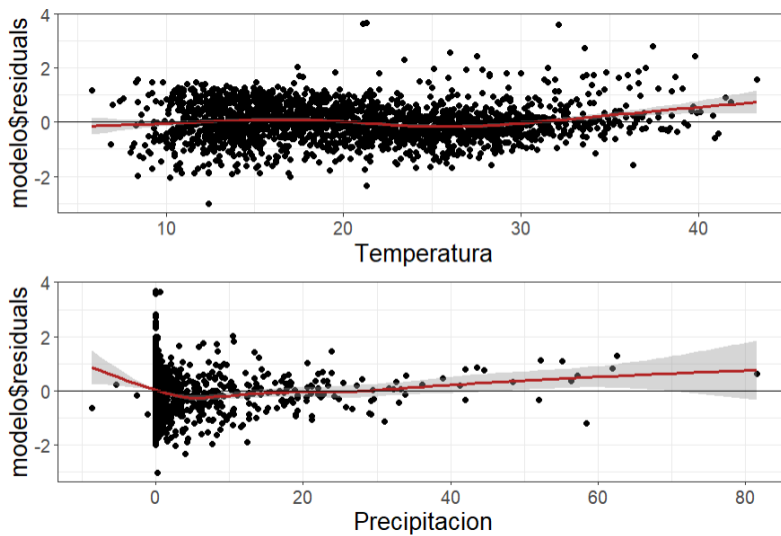


Figura 15. Correlación de la concentración promedio de PM_{10} versus la temperatura y la precipitación en Monash



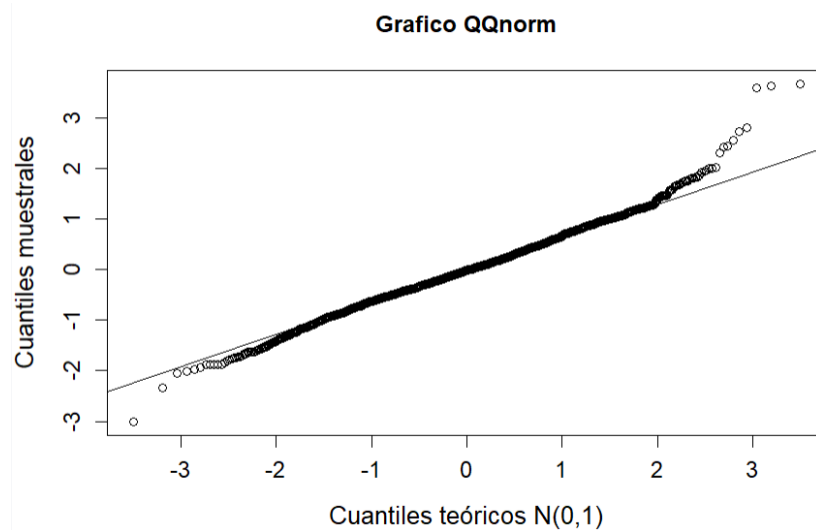
En el primer modelo general ajustado se obtuvo que la temperatura, las precipitaciones, los incendios, la fase 1, y el desconfinamiento, son variables significativas que ayudan a explicar la concentración promedio de PM_{10} en Monash (p -valor $< 0,05$). En cuanto al supuesto de linealidad, como se observa en la Figura 16, los residuos se distribuyen aleatoriamente en torno a 0, por lo que se puede afirmar que se cumple el supuesto de linealidad.

Figura 16. Comprobación del supuesto de linealidad del modelo de regresión de PM_{10} en Monash



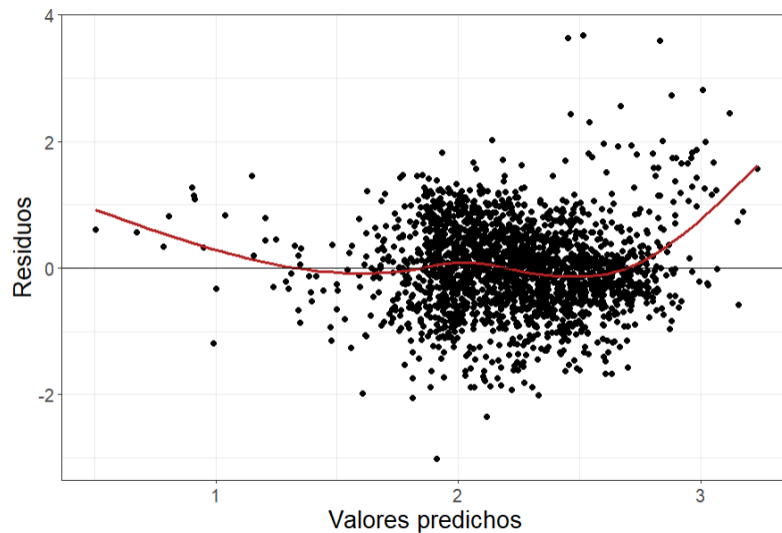
Respecto a la normalidad, en la Figura 17 se observa que la mayoría de los puntos se distribuyen sobre la diagonal, siendo el test de Kolmogorov Smirnov el que confirma que se cumple este supuesto ($p\text{-valor} = 0,07723 > 0,05$).

Figura 17. Comprobación del supuesto de normalidad del modelo de regresión de PM_{10} en Monash



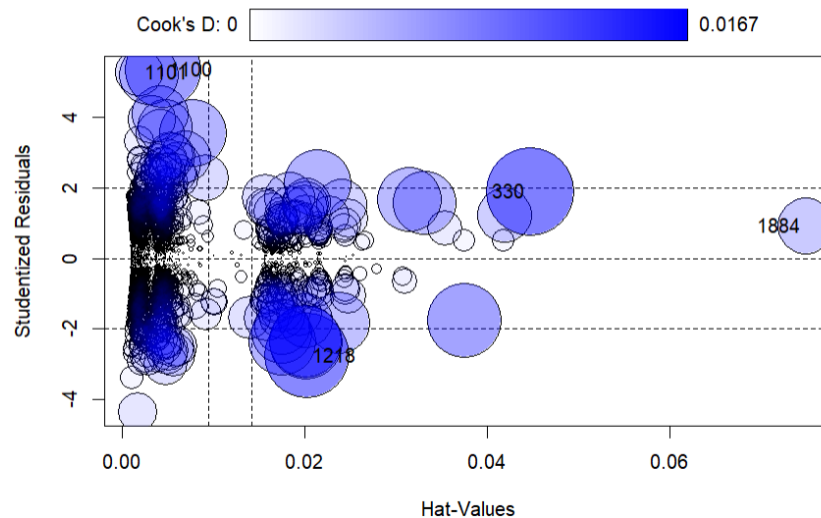
Por último, respecto a la homocedasticidad, el gráfico de variabilidad constante de los residuos (Figura 18) muestra que la dispersión de los puntos sigue un patrón similar a un óvalo, lo que sugiere que la variabilidad es dependiente del valor ajustado y por lo tanto no hay homocedasticidad, así mismo el estadístico de Breusch-Pagan nos lo confirma ($p\text{-valor} = 2,425e-11 < 0,05$).

Figura 18. Comprobación del supuesto de homocedasticidad del modelo de regresión de PM_{10} en Monash



Dado que se incumple el supuesto de homocedasticidad, es conveniente detectar los valores atípicos con la función “outlierTest” y con la función “influencePlot” los valores influyentes (Figura 19).

Figura 19. Valores influyentes en el modelo de concentración promedio de PM_{10}



Una vez excluidos de la base de datos los puntos atípicos e influyentes, se ajustó nuevamente un modelo de regresión en el método “stepwise mixto” para elegir el mejor modelo en base al indicador de Akaike (AIC). Como resultado, los resultados de la muestra son representativos de la población puesto que se obtuvo un error estándar medio del modelo ligeramente superior a 0,5 (Residual standard error: 0,6743), por otro lado, en

relación a las variables significativas y que afectan al indicador de Akaike, los coeficientes revelan que por cada grado centígrado que aumente la temperatura, las concentraciones de PM₁₀ aumentaron en un 3,23% mientras que por cada milímetro de lluvia, las concentraciones de PM₁₀ disminuyeron en un 2,49%. En relación con las variables de tipo factor, durante las etapas caracterizadas por grandes incendios forestales, se observó un aumento promedio del 23,83% en la concentración de PM₁₀. Asimismo, durante la fase 1, se registró un aumento promedio del 26,17% mientras que durante el desconfinamiento las concentraciones de PM₁₀ disminuyeron en un 29,67%.

Sin embargo, la fase 0, el confinamiento y los fines de semana no son variables significativas ($p\text{-valor} > 0,05$) ni afectan al coeficiente Akaike.

Tabla 7. Resultados del modelo de regresión para la partícula PM₁₀ en Monash

VARIABLES	COEFICIENTES	ERROR STD.	ESTADÍSTICO T	P-VALOR
INTERCEPTO	1,5399	0,0476	32,351	< 2e-16
TEMPERATURA	0,0323	0,0021	15,242	< 2e-16
PRECIPITACIÓN	-0,0249	0,0024	-10,398	< 2e-16
INCENDIO	0,2383	0,0322	7,405	1.88e-13
FASE 1	0,2617	0,0904	2,894	0.0038
DESCONFINAMIENTO	-0,2967	0,0875	-3,392	0.0007

Fuente: Elaboración propia

1.2 Resultados de los modelos de regresión en Florey

Partícula NO₂

Como parte del análisis descriptivo preliminar, la Figura 20 revela la posible presencia de valores atípicos en la concentración promedio de NO₂ en relación con cada variable predictora de tipo factor. Además, en la Figura 21 se evidencia una correlación negativa y significativa entre el promedio de concentración de NO₂ y las variables de temperatura ($r = -0,13$) y precipitaciones ($r = -0,15$), por tanto, a medida que las concentraciones de NO₂ aumentan, la temperatura tiende a disminuir y las precipitaciones

son menos frecuentes. Asimismo, en el diagrama de dispersión se pueden observar claramente algunos puntos atípicos que será conveniente excluir para ajustar mejor el modelo.

Figura 20. Diagrama de cajas de la concentración promedio de NO_2 versus las variables independientes de tipo factor en Florey

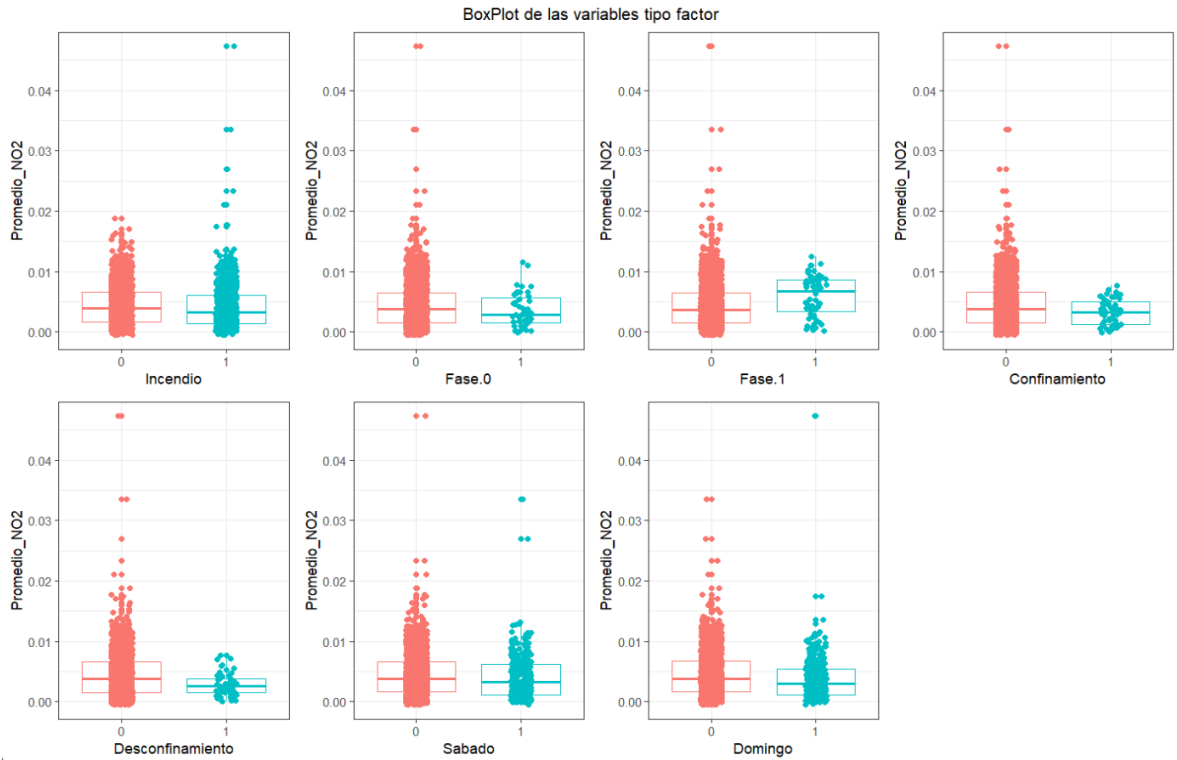
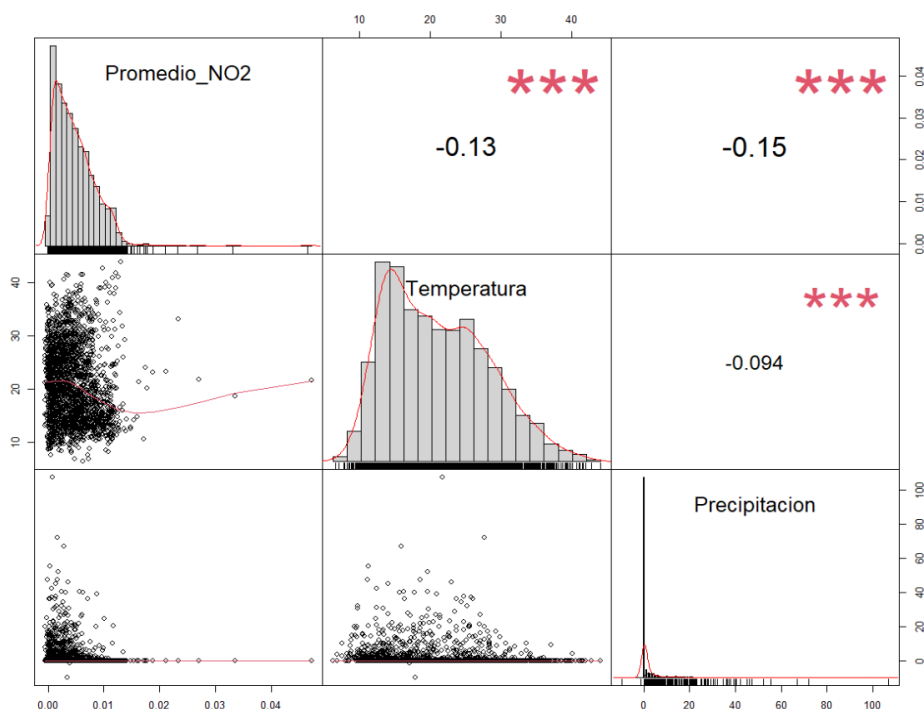


Figura 21. Correlación de la concentración promedio de NO₂ versus la temperatura y la precipitación en Florey



Con el propósito de determinar el mejor modelo de regresión, se procedió inicialmente a ajustar un modelo general. En dicho modelo, se identificó que la temperatura, las precipitaciones, los incendios, la fase 1, el confinamiento, el desconfinamiento y los fines de semana, son variables significativas ($p\text{-valor} < 0,05$) para explicar la variabilidad observada en la concentración promedio de NO₂. No obstante, como paso adicional, se realizó una evaluación de los supuestos de linealidad, normalidad y homocedasticidad.

En la Figura 22 se observa una distribución aleatoria de los residuos en torno a 0, por lo tanto, se puede afirmar que se cumple el supuesto de linealidad. Sin embargo, respecto a la normalidad, en la Figura 23 se observa que la mayoría de los puntos se distribuyen sobre la diagonal, aunque con algunos atípicos, siendo el test de Kolmogorov Smirnov el que confirma que no se cumple este supuesto ($p\text{-valor} = 3,797e-08 < 0,05$).

Figura 22. Comprobación del supuesto de linealidad del modelo de regresión de NO₂ en Florey

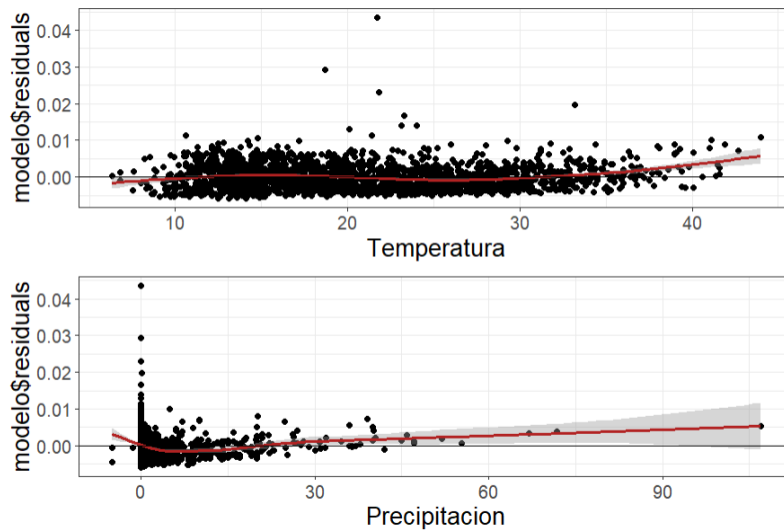
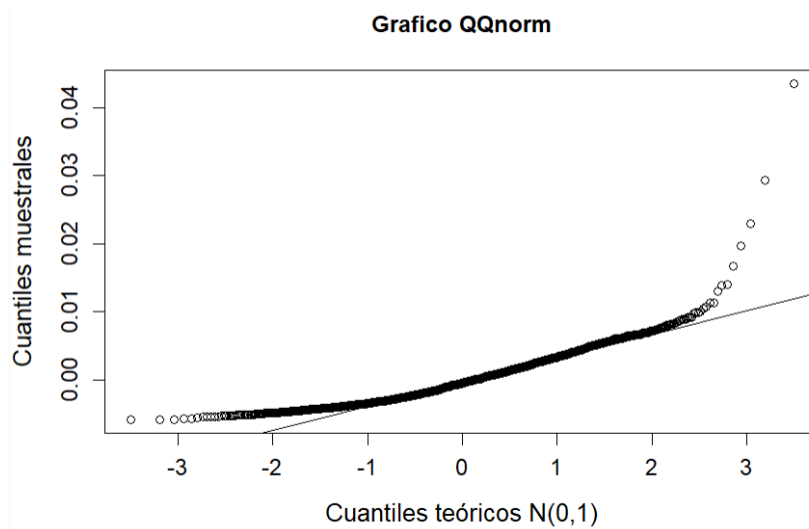
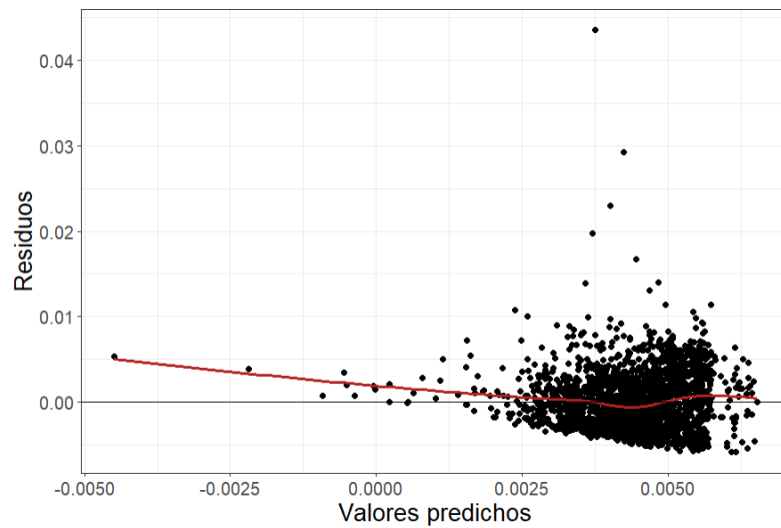


Figura 23. Comprobación del supuesto de normalidad del modelo de regresión de NO₂ en Florey



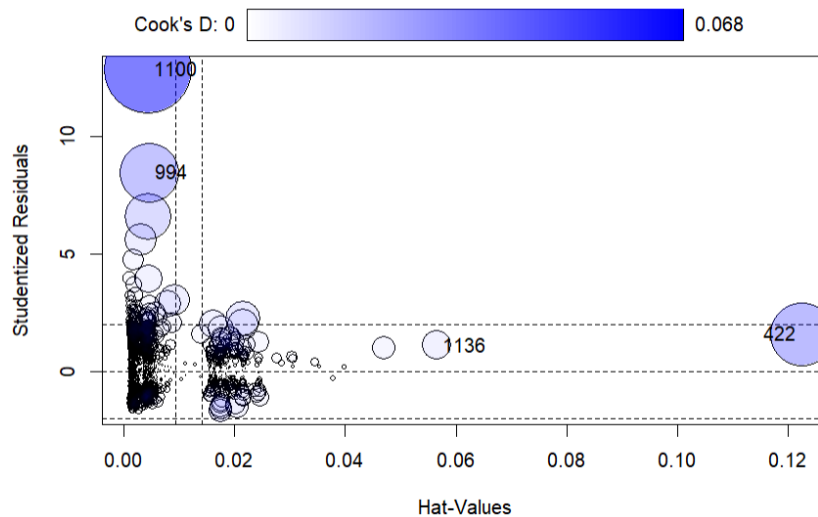
En cuanto a la homocedasticidad, el gráfico de la variabilidad constante de los residuos (Figura 24) muestra que la mayoría de los puntos no están dispersos y se concentran (se observan algunos valores dispersos), siendo el estadístico de Breusch-Pagan el que confirma que sí se cumple este supuesto ($p\text{-valor} = 0,1159 > 0,05$).

Figura 24. Comprobación del supuesto de homocedasticidad del modelo de regresión de NO₂ en Florey



Ante la evidencia de que el supuesto de normalidad no se cumple, se procedió a detectar los valores atípicos con la función “outlierTest” y con la función “influencePlot” los valores influyentes (Figura 25).

Figura 25. Valores influyentes en el modelo de concentración promedio de NO₂



Una vez excluidos de la base de datos los puntos atípicos e influyentes, se ajustó nuevamente un modelo de regresión utilizando el método “stepwise mixto” para elegir el mejor modelo en base al indicador de Akaike (AIC). Respecto al modelo final, el error estándar medio del modelo indica que los resultados de la muestra son representativos de la población dado que es menor de 0,5 (Residual standard error: 0,2183). Además, en

cuanto a las variables significativas y que influyen en el indicador de Akaike, los coeficientes indican que por cada grado centígrado que aumente la temperatura, disminuyen las concentraciones de NO₂ en un 0,48% y por cada milímetro de lluvia, las concentraciones de NO₂ disminuyen en un 0,65%. Asimismo, los incendios provocaron una disminución promedio de un 3,93% mientras que la fase 1 provocó un aumento promedio del 7,42%. El confinamiento y el desconfinamiento disminuyeron las concentraciones de NO₂ en un 10,06% y en un 9,32% respectivamente. Finalmente, los sábados y domingos la concentración de NO₂ disminuye aproximadamente un 4,72% y un 6,28% respectivamente.

Es importante mencionar que la fase 0 influye en el coeficiente de Akaike, pero no es una variable significativa (p-valor > 0,05), de hecho, durante este periodo la concentración de NO₂ disminuyó en un 6,22%.

Tabla 8. Resultados del modelo de regresión para la partícula NO₂ en Florey

VARIABLES	COEFICIENTES	ERROR STD.	ESTADÍSTICO T	P-VALOR
INTERCEPCIÓN	-4,1239	0,0160	-256,979	< 2e-16
TEMPERATURA	-0,0048	0,0007	-7,077	2.00e-12
PRECIPITACIÓN	-0,0065	0,0008	-8,096	9.54e-16
INCENDIO	-0,0393	0,0106	-3,723	0.0002
FASE 0	-0,0622	0,0324	-1,917	0.0553
FASE 1	0,0742	0,0291	2,552	0.0108
CONFINAMIENTO	-0,1006	0,0279	-3,605	0.0003
DESCONFINAMIENTO	-0,0932	0,0284	-3,278	0.0011
SÁBADO	-0,0472	0,0138	-3,412	0.0007
DOMINGO	-0,0628	0,0138	-4,557	5.50e-06

Fuente: Elaboración propia

Partícula O₃

Respecto a la concentración de O₃ en Florey, el análisis revela la posible presencia de valores atípicos en cada variable predictora de tipo factor, tal como se muestra en la

Figura 26. Además, en la Figura 27 se observa que el promedio de concentración de O_3 tiene una correlación positiva y significativa con la temperatura ($r= 0,43$) y las precipitaciones ($r= 0,065$). Por tanto, a medida que las concentraciones de O_3 aumentan, la temperatura y las precipitaciones aumentan. Asimismo, en el diagrama de dispersión se pueden identificar puntos influyentes y atípicos que será conveniente excluir para lograr un mejor ajuste del modelo.

Figura 26. Diagrama de cajas de la concentración promedio de O_3 versus las variables independientes de tipo factor en Florey

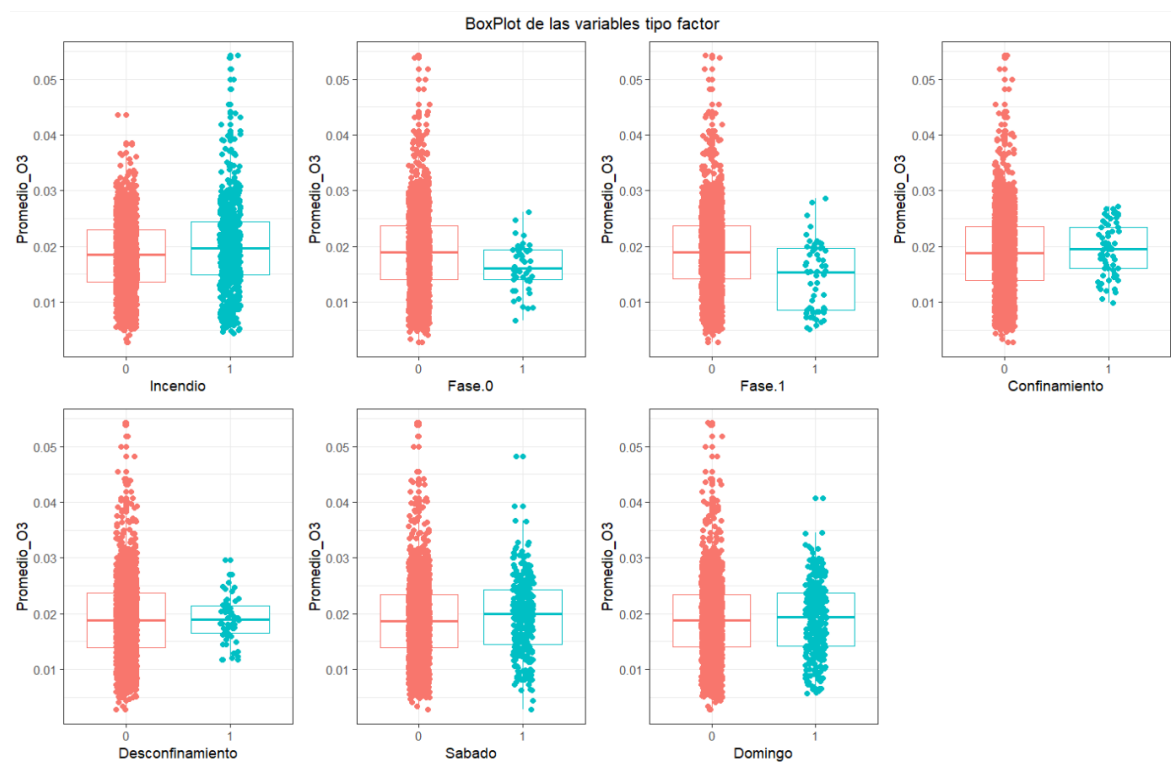
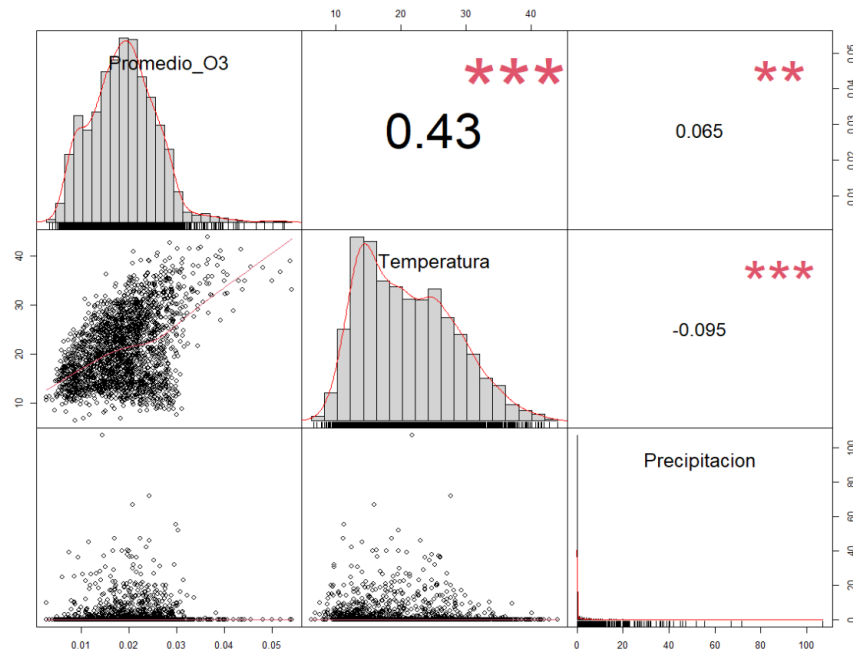


Figura 27. Correlación de la concentración promedio de O_3 versus la temperatura y la precipitación en Florey



En el primer modelo general ajustado se obtuvo que la temperatura, las precipitaciones, los incendios, la fase 0, la fase 1, el confinamiento y los sábados, son variables significativas que ayudan a explicar la concentración de O_3 (p valor $< 0,05$). Además, se realizó un análisis de los supuestos de linealidad, normalidad y homocedasticidad, en la Figura 28 se observa que los residuos se distribuyen aleatoriamente en torno a 0, lo que afirma que se cumple el supuesto de linealidad. Asimismo, respecto a la normalidad, en la Figura 29 se observan algunos valores atípicos aunque la mayoría de los puntos se distribuyen sobre la diagonal, siendo el test de Kolmogorov Smirnov el que confirma que se no cumple este supuesto (p -valor = $0,006823 < 0,05$).

Figura 28. Comprobación del supuesto de linealidad del modelo de regresión de O_3 en Florey

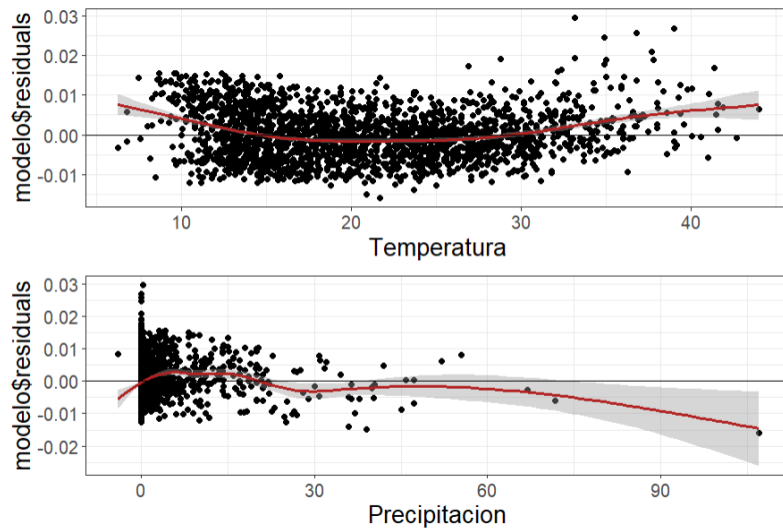
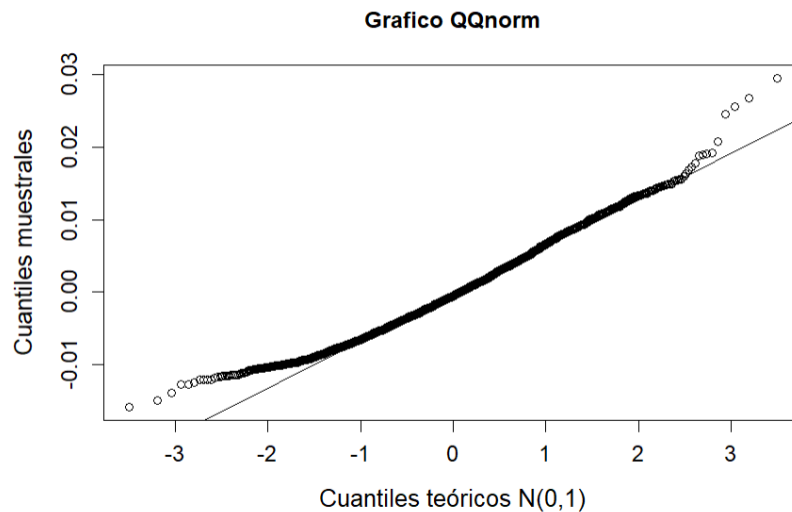
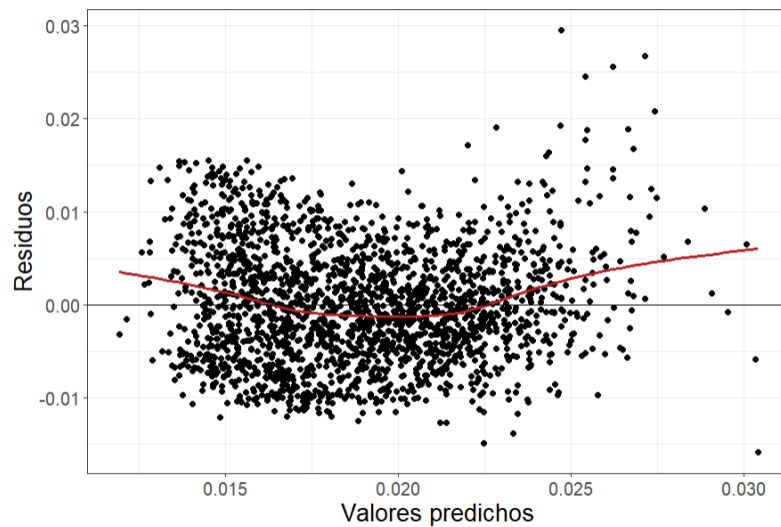


Figura 29. Comprobación del supuesto de normalidad del modelo de regresión de O_3 en Florey



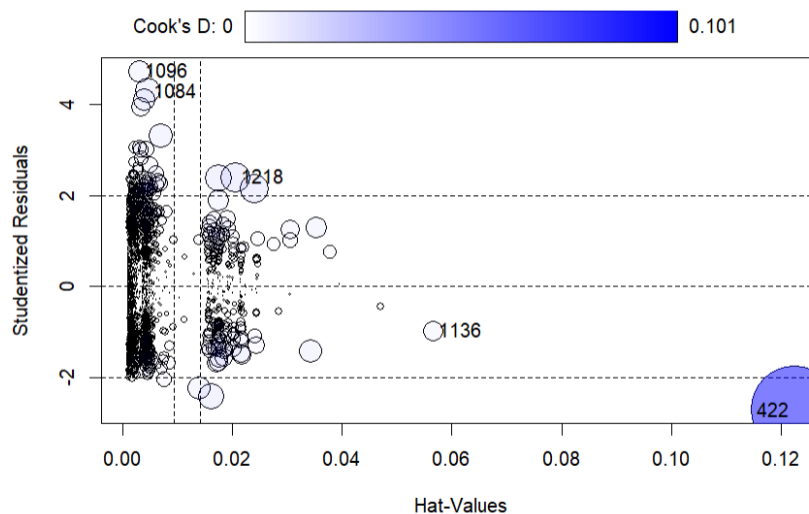
Por último, respecto a la homocedasticidad, el gráfico de variabilidad constante de los residuos (Figura 30) revela que los puntos presentan una agrupación con forma circular, lo que indica que la variabilidad es dependiente del valor ajustado y por lo tanto no hay homocedasticidad, así mismo el estadístico de Breusch-Pagan lo confirma (p -valor= $2,626e-07 < 0,05$).

Figura 30. Comprobación del supuesto de homocedasticidad del modelo de regresión de O_3 en Florey



Dado que se incumplen los supuestos de normalidad y homocedasticidad, es necesario detectar los valores atípicos con la función “outlierTest” y con la función “influencePlot” (Figura 31).

Figura 31. Valores influyentes en el modelo de concentración promedio de O_3



Una vez excluidos de la base de datos los puntos atípicos e influyentes, se ajustó nuevamente un modelo de regresión en el método “stepwise mixto” para elegir el mejor modelo en base al indicador de Akaike (AIC). En dicho modelo, se obtiene que el error estándar medio del modelo es inferior a 0,5 (Residual standard error: 0,3745), por tanto, los resultados de la muestra son representativos de la población. Asimismo, en relación a

las variables significativas y que influyen en el indicador de Akaike, los coeficientes indican que por cada grado centígrado que aumente la temperatura, las concentraciones de O₃ aumentan en un 2,43% y por cada milímetro de lluvia las O₃ aumentan aproximadamente en un 1%. Respecto a las variables de tipo factor, en el periodo que ocurrieron los grandes incendios en Australia, las concentraciones de O₃ aumentaron en un 6,07%, en la fase 1 disminuyeron en un 15,13% mientras que durante el confinamiento el promedio de concentración de O₃ aumentó en un 18,81%.

Es importante mencionar que la fase 0 y los sábados influyen en el coeficiente de Akaike, pero no son variables significativas (p-valor > 0,05), de hecho, durante la fase 0 la concentración de NO₂ disminuyó en un 8,32% mientras que los sábados aumentó un 4,52%. Por último, el desconfinamiento y los domingos ni son variables significativas ni afectan al indicador de Akaike.

Tabla 9. Resultados del modelo de regresión para la partícula O₃ en Florey

Variab les	C oeficientes	E rror std.	E stadístico t	P -valor
INTERCEPTO	-4,5988	0,0272	-169,273	< 2e-16
TEMPERATURA	0,0243	0,0012	20,742	< 2e-16
PRECIPITACIÓN	0,0090	0,0014	6,618	4.60e-11
INCENDIO	0,0607	0,0180	3,376	0.0007
FASE 0	-0,0832	0,0556	-1,497	0.1346
FASE 1	-0,1513	0,0503	-3,007	0.0027
CONFINAMIENTO	0,1881	0,0478	3,934	8.64e-05
SÁBADO	0,0452	0,0234	1,935	0.0531

Fuente: Elaboración propia

Partícula PM₁₀

Respecto a la concentración de partículas de PM₁₀ en Florey, se observa en la Figura 32 la posible existencia de valores atípicos por cada variable predictora de tipo factor, así mismo en la Figura 33 se observa que el promedio de concentración de PM₁₀ tiene una correlación positiva y significativa con la temperatura (r= 0,21) y una

correlación negativa y significativa con las precipitaciones ($r = -0,072$), lo que indica que a medida que las concentraciones de PM_{10} aumentan, la temperatura aumenta y las precipitaciones disminuyen, por otro lado, en el diagrama de dispersión, se observan claramente puntos influyentes y valores atípicos que será conveniente excluir para lograr un mejor ajuste del modelo.

Figura 32. Diagrama de cajas de la concentración promedio de PM_{10} versus las variables independientes de tipo factor en Florey

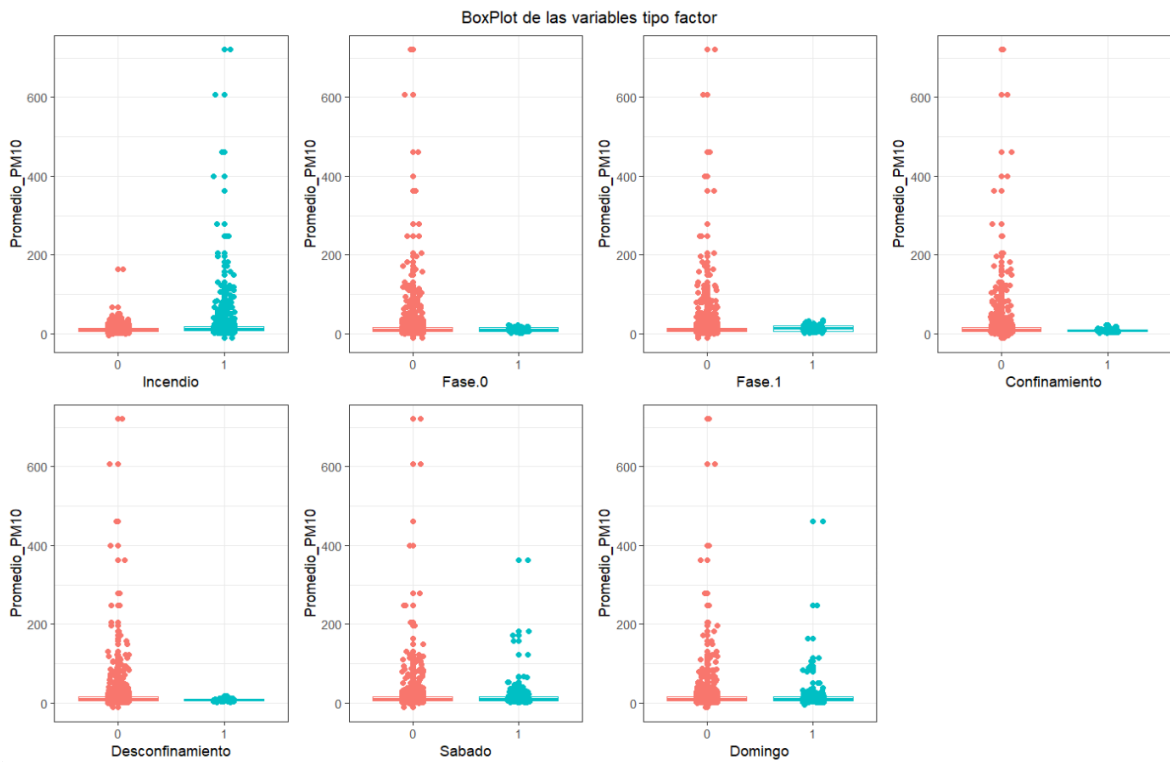
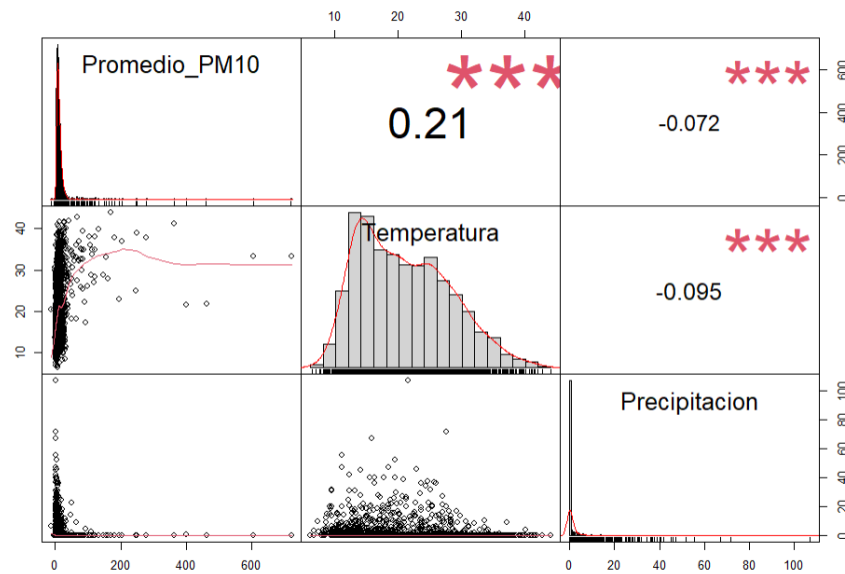


Figura 33. Correlación de la concentración promedio de PM_{10} versus la temperatura y la precipitación en Florey



En el primer modelo general ajustado se obtuvo que la temperatura, las precipitaciones y los incendios, son variables significativas que nos ayudan a explicar la concentración promedio de PM_{10} en Florey (p -valor $< 0,05$). Respecto a los supuestos de linealidad, normalidad y homocedasticidad, tal como se observa en la Figura 34, los residuos se distribuyen aleatoriamente en torno a 0, por lo que se puede afirmar que se cumple el supuesto de linealidad, en cuanto a la normalidad, en la Figura 35 se observa que la mayoría de los puntos se distribuyen sobre la diagonal (se observan algunos valores atípicos), siendo el test de Kolmogorov Smirnov el que confirma que no se cumple este supuesto (p -valor = $2.2e-16 < 0,05$). Por último, respecto a la homocedasticidad, el gráfico de variabilidad constante de los residuos (Figura 36) revela que la dispersión de los puntos se agrupa en una recta, por tanto, la variabilidad es dependiente del valor ajustado y no hay homocedasticidad, así mismo el estadístico de Breusch-Pagan lo confirma (p -valor = $0,0006474 < 0,05$).

Figura 34. Comprobación del supuesto de linealidad del modelo de regresión de PM_{10} en Florey

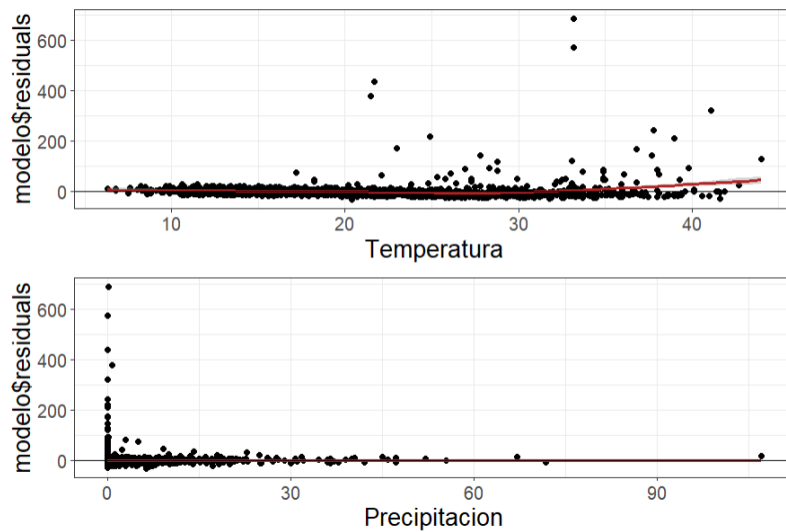


Figura 35. Comprobación del supuesto de normalidad del modelo de regresión de PM_{10} en Florey

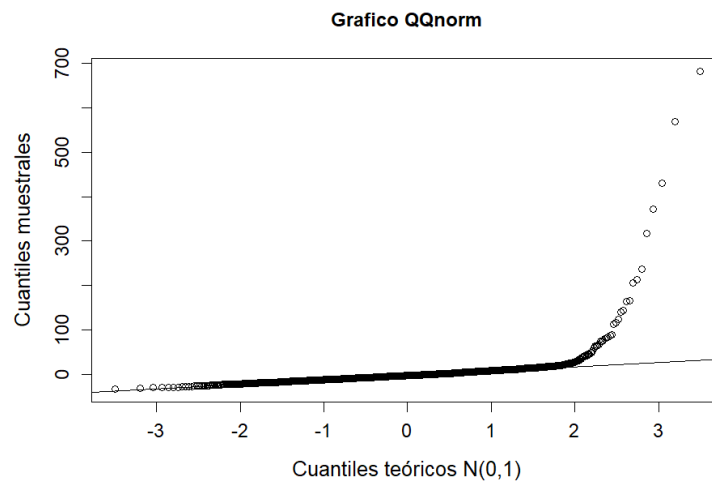
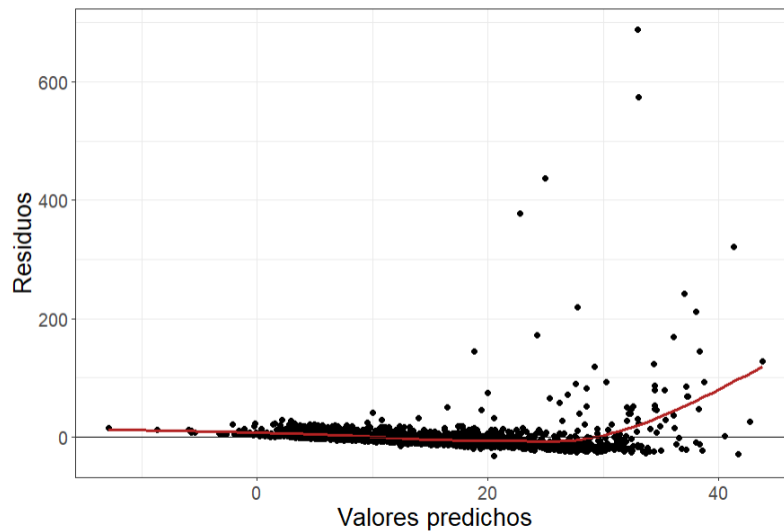
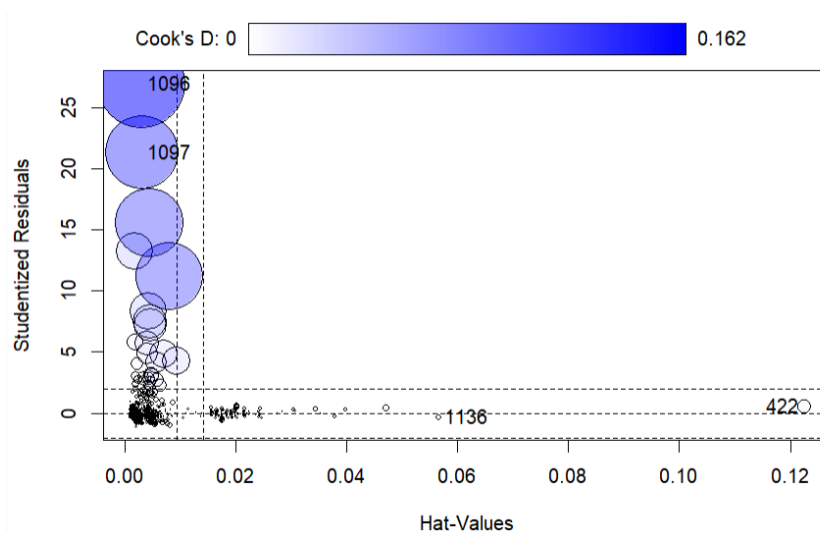


Figura 36. Comprobación del supuesto de homocedasticidad del modelo de regresión de PM_{10} en Florey



Dado que se no se cumplen los supuestos de normalidad y homocedasticidad, es conveniente detectar los valores atípicos con la función “outlierTest” y con la función “influencePlot” los valores influyentes (Figura 37).

Figura 37. Valores influyentes en el modelo de concentración promedio de PM_{10}



Una vez excluidos de la base de datos los puntos atípicos e influyentes, se ajustó nuevamente un modelo de regresión con el método “stepwise mixto” para elegir el mejor modelo en base al indicador de Akaike (AIC). Los resultados de la muestra son representativos de la población puesto que el error estándar medio del modelo resultó

ligeramente superior a 0,5 (Residual standard error: 0,6446). Además, en cuanto a las variables significativas y que influyen en el coeficiente de Akaike, los coeficientes indican por cada grado centígrado que aumente la temperatura, aumentan las concentraciones de PM₁₀ en un 2,89% y por cada milímetro de lluvia, disminuyen las concentraciones de PM₁₀ en un 2,30%. Respecto a las variables de tipo factor, en los periodos en los que ocurrieron los incendios, el promedio de concentración de PM₁₀ aumentó en un 26,71%, así mismo en la fase 1, el promedio de concentración de PM₁₀ aumentó en un 24,46%.

Por otro lado, las variables que no son significativas (p-valor > 0,05) para el modelo de regresión, pero que influyen en el indicador de Akaike fueron la fase 0 y los sábados, los coeficientes muestran que durante la fase 0 la concentración de PM₁₀ aumentó un 7,20% y los sábados un 6,08%. Por último, el confinamiento, el desconfinamiento y los domingos no son variables significativas ni afectan al indicador de Akaike.

Tabla 10. Resultados del modelo de regresión para la partícula PM₁₀ en Florey

Variables	Coefficientes	Error std.	Estadístico t	P-valor
INTERCEPTO	1,6141	0,0462	34,917	< 2e-16
TEMPERATURA	0,0289	0,0020	14,321	< 2e-16
PRECIPITACIÓN	-0,0230	0,0022	-10,477	< 2e-16
INCENDIO	0,2671	0,0309	8,652	< 2e-16
FASE 0	0,0720	0,0956	0,753	0.4516
FASE 1	0,2446	0,0858	2,85	0.0044
SÁBADO	0,0608	0,0402	1,512	0.1308

Fuente: Elaboración propia

1.3 Resultados de los modelos de regresión en Civic

Partícula O₃

En el análisis descriptivo previo, la Figura 38 muestra la posible existencia de valores atípicos de la concentración promedio de O₃ por cada variable predictora de tipo factor, así mismo en la Figura 39 se observa que el promedio de concentración de O₃ tiene una correlación positiva y significativa con la temperatura ($r= 0,5$) mientras que con las precipitaciones ($r= 0,016$) la correlación es positiva pero no es significativa, no obstante, en el modelo final sí que será una variable significativa, por tanto, a medida que las concentraciones de O₃ aumentan, la temperatura y las precipitaciones aumentan, por otro lado, en el diagrama de dispersión, se observan puntos atípicos que será conveniente excluir para lograr un mejor ajuste del modelo.

Figura 38. Diagrama de cajas de la concentración promedio de O₃ versus las variables independientes de tipo factor en Civic

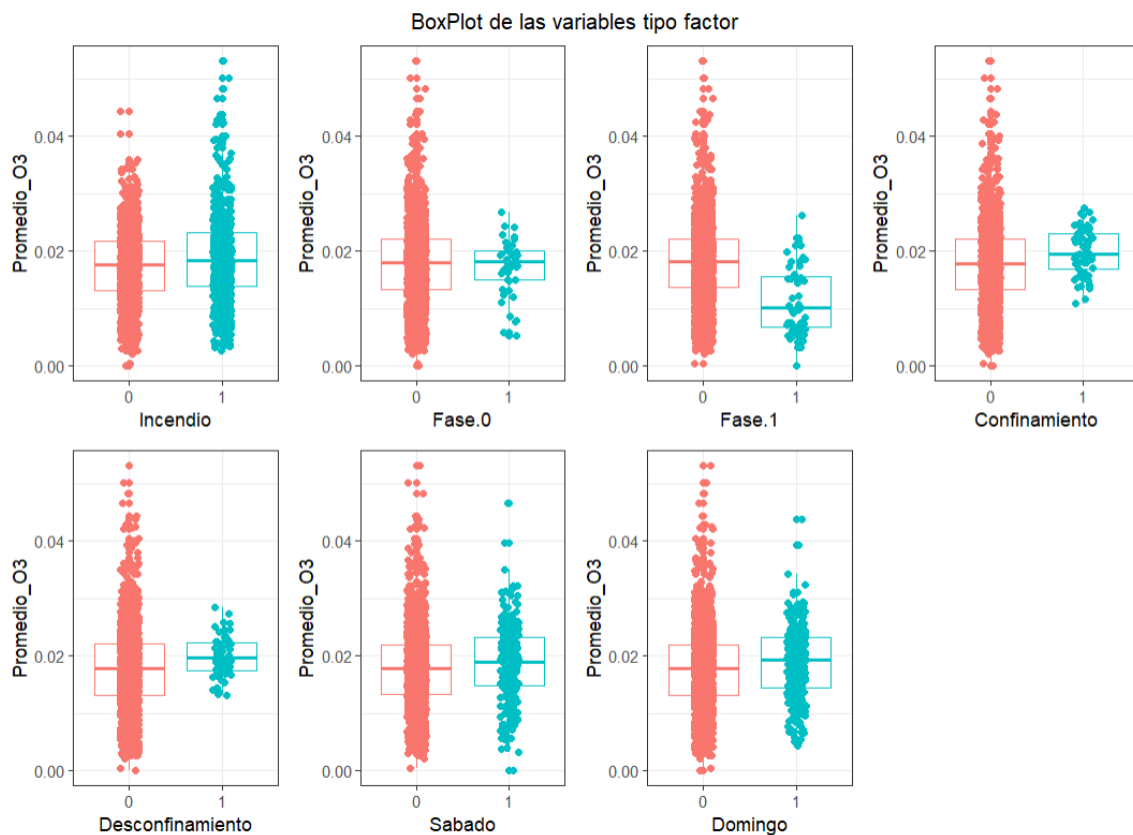
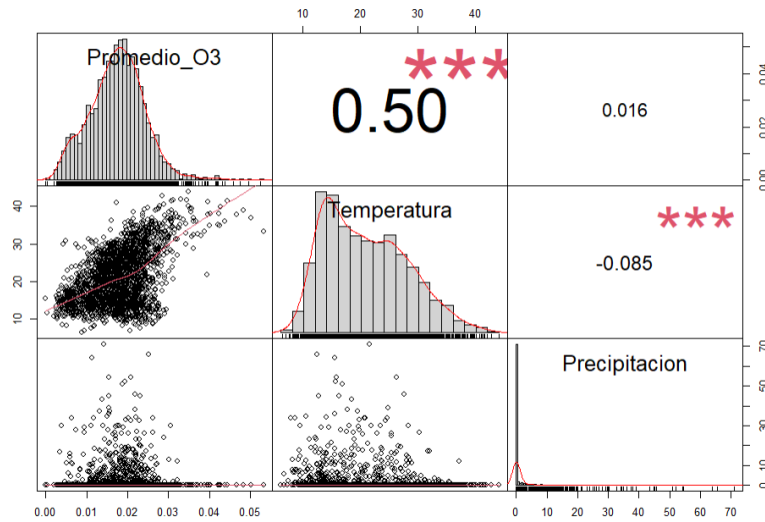


Figura 39. Correlación de la concentración promedio de O_3 versus la temperatura y la precipitación en Civic



En el primer modelo general ajustado se obtuvo que la temperatura, las precipitaciones, los incendios, la fase 1, el confinamiento, el desconfinamiento y los fines de semana, son variables significativas para explicar la concentración de O_3 en Civic (p -valor $< 0,05$). En cuanto a los supuestos de linealidad, normalidad y homocedasticidad, tal como se observa en la Figura 40, los residuos se distribuyen aleatoriamente en torno a 0, por lo que se cumple el supuesto de linealidad, respecto a la normalidad, en la Figura 41 se observa que la mayoría de los puntos se distribuyen sobre la diagonal, aunque con algunos atípicos, siendo el test de Kolmogorov Smirnov el que confirma que se cumple este supuesto (p -valor = $0,07009 > 0,05$).

Figura 40. Comprobación del supuesto de linealidad del modelo de regresión de O_3 en Civic

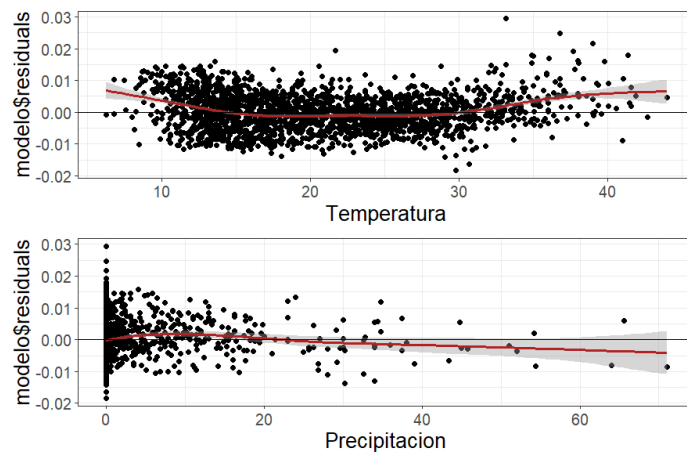
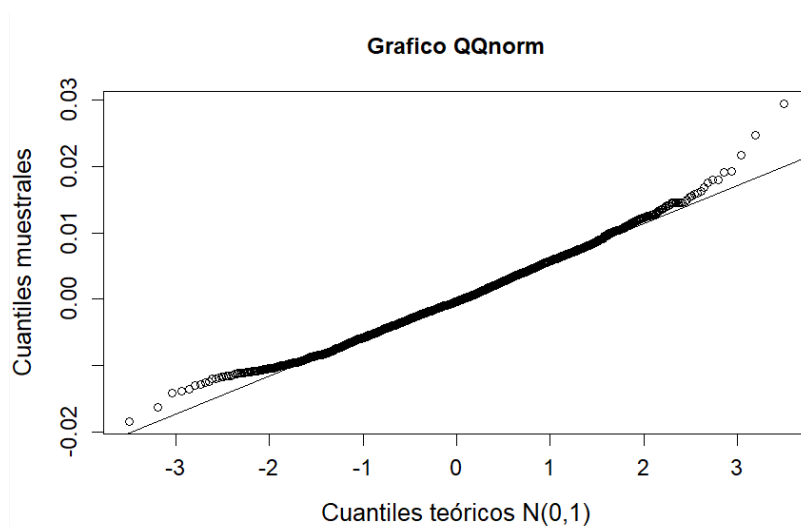
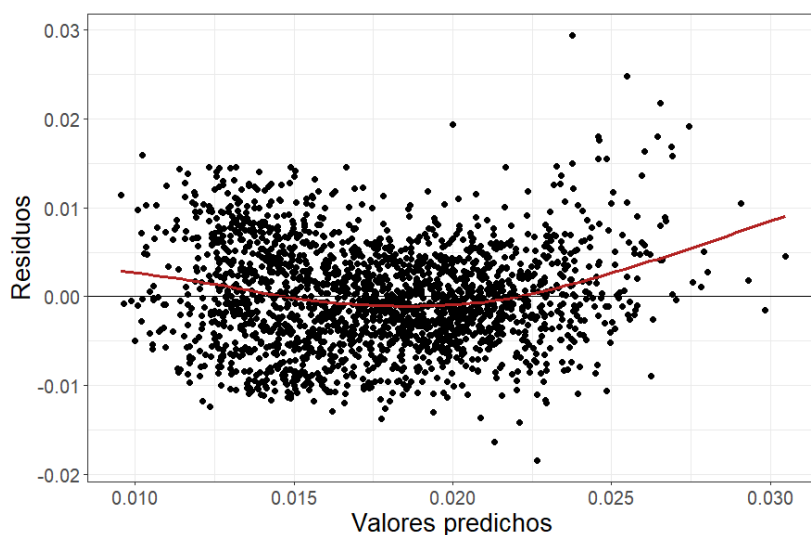


Figura 41. Comprobación del supuesto de normalidad del modelo de regresión de O_3 en Civic



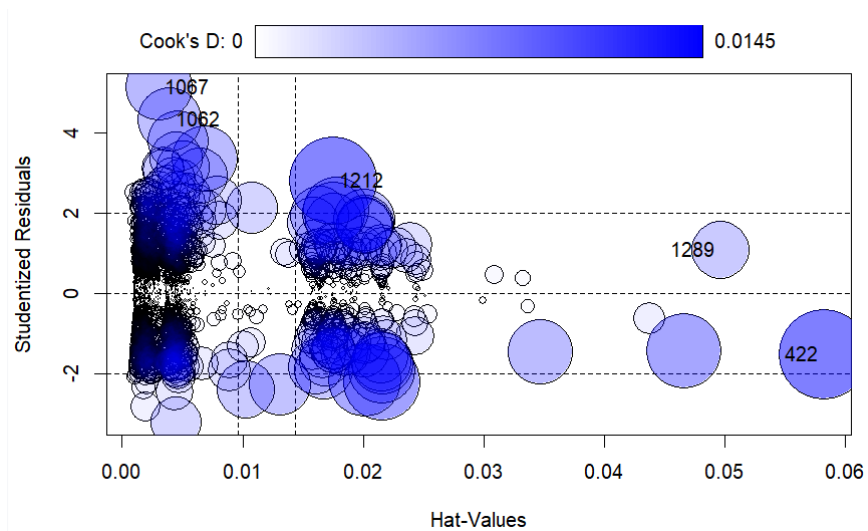
Finalmente, respecto a la homocedasticidad, el gráfico de variabilidad constante de los residuos (Figura 42) muestra que los puntos se agrupan formando una distribución circular, como consecuencia, la variabilidad es dependiente del valor ajustado y por lo tanto no hay homocedasticidad, así mismo el estadístico de Breusch-Pagan lo confirma ($p\text{-valor} = 0,0001183 < 0,05$).

Figura 42. Comprobación del supuesto de homocedasticidad del modelo de regresión de O_3 en Civic



Puesto que se incumple el supuesto de homocedasticidad, es necesario detectar los valores atípicos con la función “outlierTest” y con la función “influencePlot” los valores influyentes (Figura 43).

Figura 43. Valores influyentes en el modelo de concentración promedio de O_3



Una vez excluidos de la base de datos los puntos atípicos e influyentes, se ajustó nuevamente un modelo de regresión en el método “stepwise mixto” para elegir el mejor modelo en base al indicador de Akaike (AIC). En el nuevo modelo, los resultados de la muestra son representativos de la población puesto que se obtuvo un error estándar medio del modelo inferior a 0,1 (Residual standard error: 0,04833). En primer lugar, respecto a las variables significativas y que afectan al indicador de Akaike, los coeficientes indican que la temperatura, las precipitaciones y los incendios de gran escala ocurridos en Australia han influido en el incremento de concentración de O_3 , ya que por cada grado centígrado que aumente la temperatura, las concentraciones de O_3 aumentaron en un 0,40%, y por cada milímetro de lluvia aumentaron en un 0,07%, y a causa de los incendios aumentaron en un 0,93%. Asimismo, durante la fase 1 disminuyeron las concentraciones en un 3,13% mientras que durante el confinamiento y el desconfinamiento, el promedio de concentración de O_3 aumentó en un 3,85% y en un 2,35% respectivamente. Finalmente, las concentraciones de O_3 aumentaron en un 1,36% y en un 1,56% los sábados y los domingos.

Por último, la fase 0 no es una variable significativa ($p\text{-valor} > 0,05$) ni influye en el coeficiente Akaike.

Tabla 11. Resultados del modelo de regresión para la partícula O₃ en Civic

Variab les	C oeficientes	E rror std.	E stadístico t	P -valor
INTERCEPTO	-2,2334	0,0036	-627,624	<2e-16
TEMPERATURA	0,0040	0,0002	26,238	<2e-16
PRECIPITACIÓN	0,0007	0,0002	4,178	3,07e-05
INCENDIO	0,0093	0,0023	4,012	6,23e-05
FASE 1	-0,0313	0,0065	-4,815	1,58e-06
CONFINAMIENTO	0,0385	0,0062	6,239	5,33e-10
DESCONFINAMIENTO	0,0235	0,0063	3,734	0,0002
SÁBADO	0,0136	0,0031	4,407	1,10e-05
DOMINGO	0,0156	0,0031	5,065	4,44e-07

Fuente: Elaboración propia

Partícula PM₁₀

En relación a la concentración de partículas de PM₁₀ en Civic, en la Figura 44 se puede apreciar la posible existencia de valores atípicos por cada variable predictora de tipo factor, así mismo en la Figura 45 se observa que el promedio de concentración de PM₁₀ tiene una correlación positiva y significativa con la temperatura ($r= 0,24$) mientras que con las precipitaciones ($r= 0,025$) la correlación es positiva pero no es significativa, no obstante, en el modelo final sí que será una variable significativa, por tanto, a medida que las concentraciones de PM₁₀ aumentan, la temperatura y las precipitaciones aumentan, por otro lado, en el diagrama de dispersión, se pueden identificar puntos influyentes y valores atípicos que será conveniente excluir con el objetivo de mejorar el modelo.

Figura 44. Diagrama de cajas de la concentración promedio de PM_{10} versus las variables independientes de tipo factor en Civic

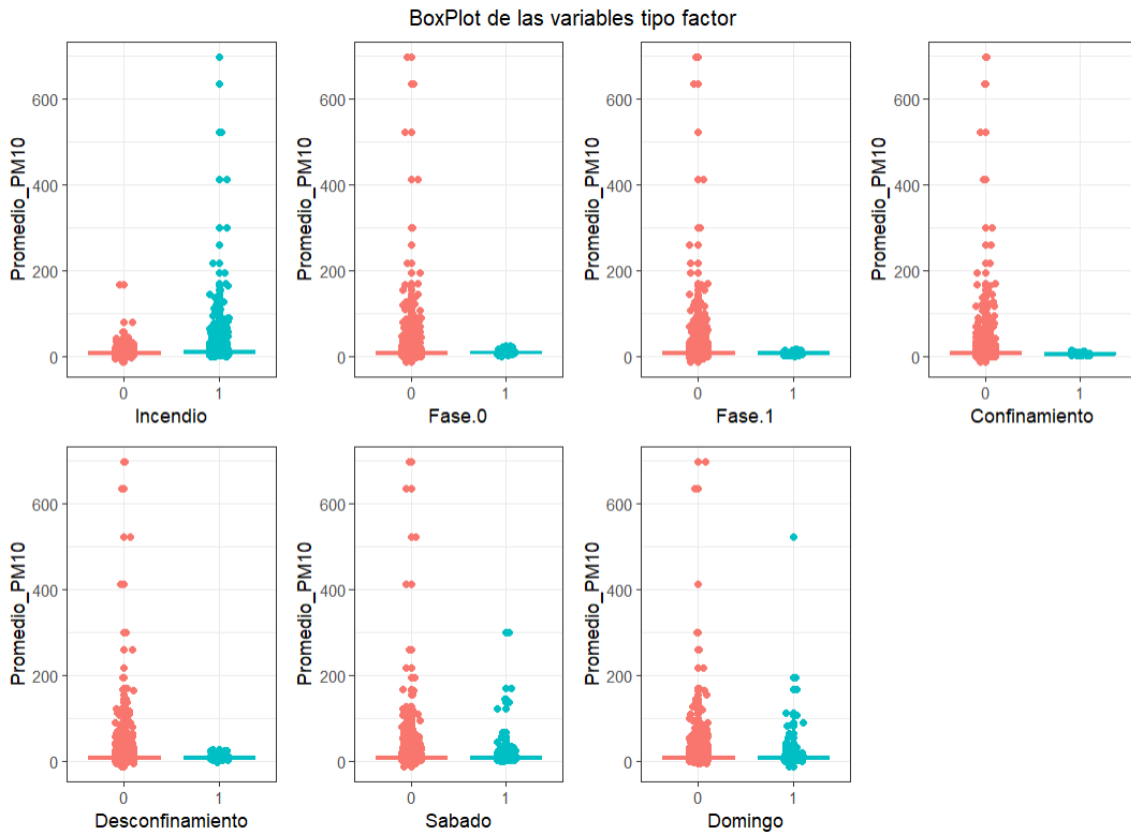
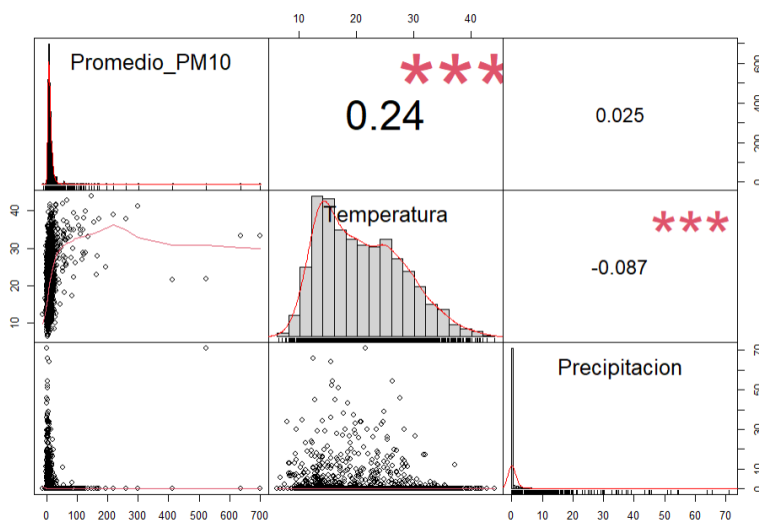


Figura 45. Correlación de la concentración promedio de PM_{10} versus la temperatura y la precipitación en Civic



En el primer modelo general ajustado se obtuvo que la temperatura, las precipitaciones y los incendios, son variables significativas que permiten explicar la concentración promedio de PM_{10} en Civic (p -valor < 0,05). En cuanto a los supuestos de

linealidad, normalidad y homocedasticidad, tal como se observa en la Figura 46, los residuos se distribuyen aleatoriamente en torno a 0, lo que confirma que se cumple el supuesto de linealidad, respecto a la normalidad, en la Figura 47 se observa que la mayoría de los puntos se distribuyen sobre la diagonal, siendo el test de Kolmogorov Smirnov el que confirma que no se cumple este supuesto ($p\text{-valor} = 2.2e-16 < 0,05$).

Figura 46. Comprobación del supuesto de linealidad del modelo de regresión de PM_{10} en Civic

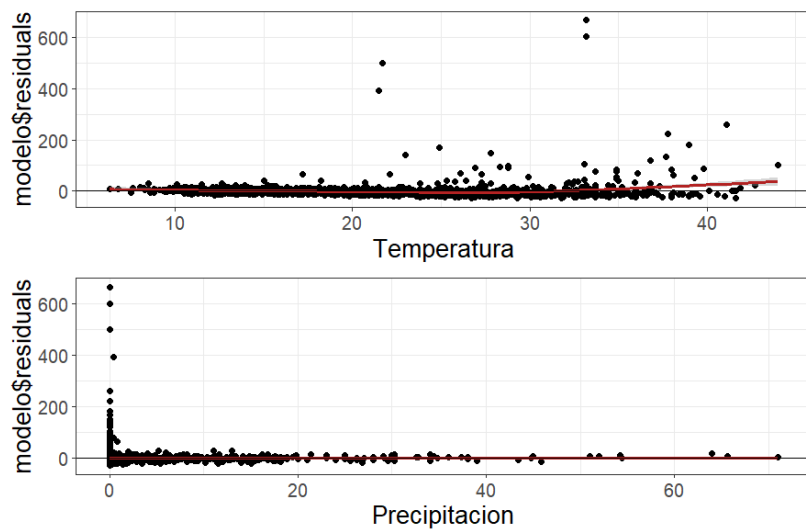
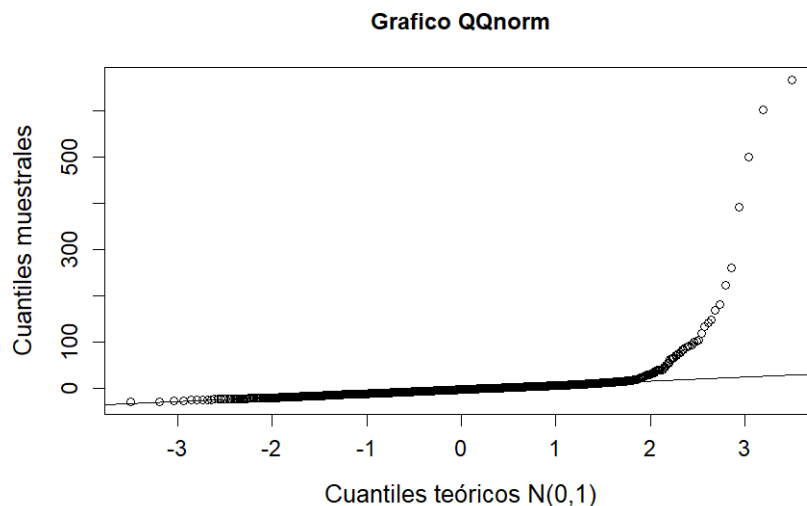


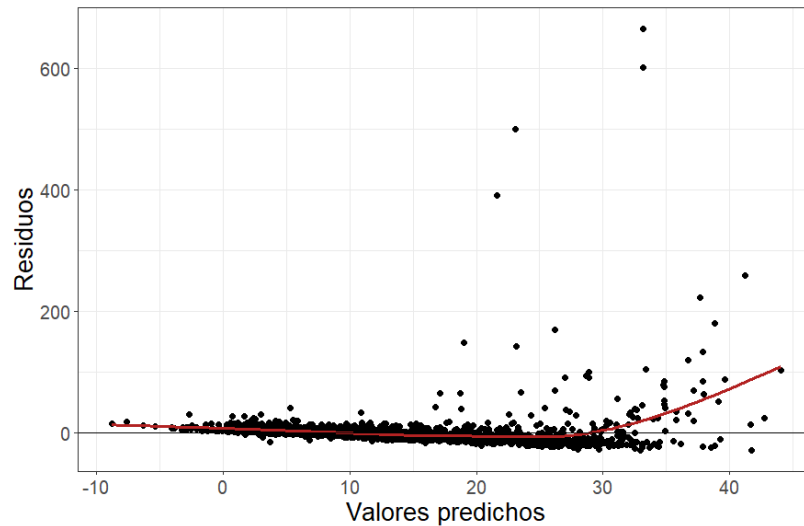
Figura 47. Comprobación del supuesto de normalidad del modelo de regresión de PM_{10} en Civic



Por último, respecto a la homocedasticidad, el gráfico de variabilidad constante de los residuos (Figura 48) muestra que los puntos se agrupan en una recta, por tanto, la

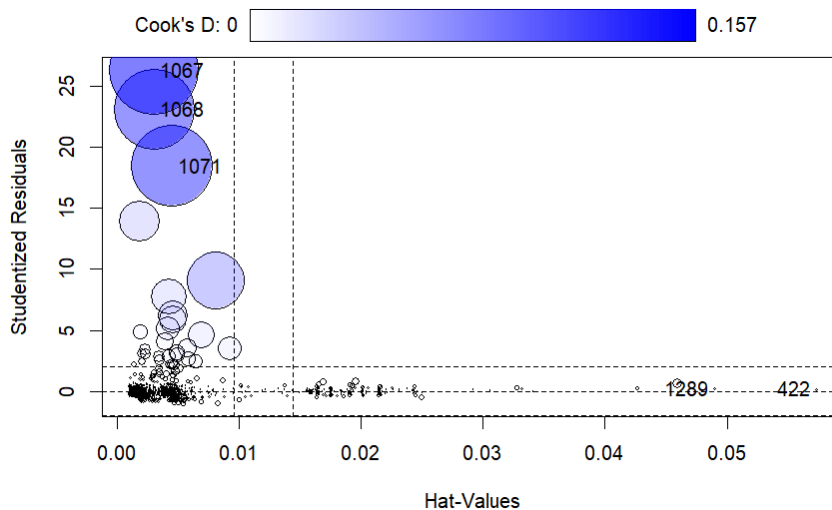
variabilidad es dependiente del valor ajustado y no hay homocedasticidad, además de que el estadístico de Breusch-Pagan lo confirma ($p\text{-valor} = 0,003284 < 0,05$).

Figura 48. Comprobación del supuesto de homocedasticidad del modelo de regresión de PM_{10} en Civic



Dado que se incumplen los supuestos de normalidad y homocedasticidad, es conveniente detectar los valores atípicos con la función “outlierTest” y con la función “influencePlot” los valores influyentes (Figura 49).

Figura 49. Valores influyentes en el modelo de concentración promedio de PM_{10}



Una vez excluidos de la base de datos los puntos atípicos e influyentes, se ajustó nuevamente un modelo de regresión en el método “stepwise mixto” para elegir el mejor modelo en base al indicador de Akaike (AIC). El error estándar medio del nuevo modelo

fue ligeramente superior a 0,5 (Residual standard error: 0,5944), por tanto, los resultados de la muestra son representativos de la población. Por otro lado, respecto a las variables significativas y que afectan al indicador de Akaike, los coeficientes indican por cada grado centígrado que aumente la temperatura, aumentan las concentraciones de PM₁₀ en un 3,97% y por cada milímetro de lluvia, disminuyen las concentraciones de PM₁₀ en un 1,79%. Respecto a las variables de tipo factor y de acuerdo al modelo generado, en las etapas donde hubo grandes incendios, el promedio de concentración de PM₁₀ aumentó en un 17,11%.

Finalmente, las variables que no son significativas ni influyen en el indicador Akaike son: la fase 0, la fase 1, el confinamiento, el desconfinamiento y los fines de semana.

Tabla 12. Resultados del modelo de regresión para la partícula PM₁₀ en Civic

Variables	Coefficientes	Error std.	Estadístico t	P-valor
INTERCEPTO	1,4081	0,0414	34,008	<2e-16
TEMPERATURA	0,0397	0,0018	21,598	<2e-16
PRECIPITACIÓN	-0,0179	0,0021	-8,603	<2e-16
INCENDIO	0,1712	0,0283	6,045	1,76e-09

Fuente: Elaboración propia

Tabla 13. Resultados de la regresión por contaminante en Canberra

Variable independiente	Descripción	NO₂	O₃	PM₁₀
Temperatura máxima (C°)	Valor medio	-0,0072	0,0168	0,0337
	Negativo y significativo (%)	100%	0%	0%
	Positivo y significativo (%)	0%	100%	100%
Precipitación (mm)	Valor medio	-0,0073	0,0060	-0,0219
	Negativo y significativo (%)	100%	0%	100%
	Positivo y significativo (%)	0%	100%	0%
Incendio (1:Si, 0:No)	Valor medio	-0,0335	0,0350	0,2255
	Negativo y significativo (%)	100%	0%	0%
	Positivo y significativo (%)	0%	100%	100%
Fase 0 (1:Si, 0:No)	Valor medio	-0,0738	-0,0832	0,0720
	Negativo y significativo (%)	50%	0%	0%
	Positivo y significativo (%)	0%	0%	0%
Fase 1 (1:Si, 0:No)	Valor medio	0,0880	-0,1751	0,2531
	Negativo y significativo (%)	100%	100%	0%
	Positivo y significativo (%)	0%	0%	100%
Confinamiento (1:Si, 0:No)	Valor medio	-0,0719	0,1255	
	Negativo y significativo (%)	50%	0%	
	Positivo y significativo (%)	0%	100%	
Desconfinamiento (1:Si, 0:No)	Valor medio	-0,1328	0,0589	-0,2967
	Negativo y significativo (%)	100%	0%	100%
	Positivo y significativo (%)	0%	50%	0%
Sabado (1:Si, 0:No)	Valor medio	-0,0448	0,0366	0,0608
	Negativo y significativo (%)	100%	0%	0%
	Positivo y significativo (%)	0%	100%	0%
Domingo (1:Si, 0:No)	Valor medio	-0,0595	0,0156	
	Negativo y significativo (%)	100%	0%	
	Positivo y significativo (%)	0%	100%	
Error estándar	Valor medio	0, 2234	0,2678	0,6377

Fuente: Elaboración propia

2. DISCUSIÓN DE LOS RESULTADOS OBTENIDOS

El presente trabajo se ha centrado en el estudio de los niveles de contaminación de Canberra en el sexenio 2017-2022 y en concreto en la realización de un análisis de aquellas variables que podían repercutir en dichos niveles de concentración de contaminantes en el sexenio indicado. La contaminación se ha convertido en una cuestión de gran relevancia y preocupación para la sociedad debido a su impacto en la salud y el medio ambiente. Esta relevancia se refleja en el interés de numerosos expertos que

analizan la contaminación en distintas ciudades y países del mundo. Por ejemplo, Molina et al. (2007) realizan un estudio de la calidad del aire en la ciudad de México, Titl (2019) estudia la crisis por la contaminación atmosférica en China, Sharma y Mauzerall (2022) llevan a cabo un estudio sobre la calidad del aire en India y así multitud de ejemplos en diversidad de localizaciones sobre la contaminación ambiental.

Además, la investigación sobre los niveles de contaminación adquiere un mayor interés a raíz de las restricciones de movilidad impuestas como consecuencia del COVID-19. Las restricciones impuestas en multitud de países para frenar la pandemia supusieron una reducción de la emisión antropogénica de contaminantes. Es por ello, que estas restricciones suponen una oportunidad única para estudiar la reducción de la contaminación ante la reducción de la emisión de contaminantes por parte de los humanos (Albayati, 2021).

En el caso de Australia, y concretamente Canberra, resulta interesante el estudio de la contaminación debido al alto número de incendios que se originan en los alrededores de la ciudad. Canberra es una ciudad planificada y con un diseño que permite que sus niveles de contaminación sean inferiores en comparación a otras grandes ciudades. Sin embargo, debido a los incendios, Canberra ha llegado a ocupar el puesto 49 de las capitales más contaminadas (IQAir, s.f.)

Del estudio de los datos obtenidos por tres estaciones de medición localizadas en Canberra (Civic, Monash y Florey), se han obtenido una serie de conclusiones sobre ciertos factores preseleccionados que afectan a la concentración de las tres partículas que se han seleccionado (NO_2 , PM_{10} y O_3). Los resultados obtenidos muestran los patrones en la concentración de contaminantes ante la variación de los siguientes factores: (i) Temperatura máxima, (ii) Precipitación, (iii) Incendios, (iv) Fase 0, (v) Fase 1, (vi) Confinamiento, (vii) Desconfinamiento, (viii) Sábado, y (ix) Domingo.

Estos factores solo serán objeto de estudio cuando sean variables significativas ($p\text{-valor} < 0,05$) que permitan explicar la concentración promedio de los contaminantes.

En cuanto al contaminante NO_2 , se han obtenido resultados de las estaciones Monash y Florey. El dióxido de nitrógeno (NO_2) es un gas producido principalmente por

la quema de combustibles fósiles, como en los motores de vehículos y en la generación de energía. Su formación se debe a la oxidación del óxido de nitrógeno (NO) en presencia de oxígeno en la atmósfera (Kumar, 2010).

En primer lugar, la concentración de NO₂ ante un aumento de temperatura se reduce en Monash un 0,96% y en Florey un 0,48% por cada grado centígrado. Esto se debe a que las reacciones químicas que forman el NO₂ tienden a ser más lentas a temperaturas más altas. Además, las altas temperaturas favorecen la dispersión y dilución del NO₂ en el aire, lo que reduce su concentración (Martínez, 2021). En cuanto a las precipitaciones, los coeficientes muestran que la lluvia también contribuye a la disminución de las concentraciones de NO₂ en ambas estaciones con una reducción de un 0,82% en Monash por cada milímetro de lluvia y de un 0,65% en Florey, ya que las gotas de lluvia arrastran y atrapan el NO₂ hasta depositarlo en el suelo (Pabón et al., s.f., p. 50).

En cuanto al periodo relativo al COVID-19, los resultados indican que, durante el primer periodo de restricciones, el cual se ha denominado fase 0, la concentración de NO₂ disminuyó tanto en la estación Monash como en Florey un 8,54% y un 6,22% respectivamente (en el caso de Florey la fase 0 no es una variable significativa pero afecta al indicador de Akaike). Emmerson y Keyword, (2021) explican el caso de la ciudad australiana de Brisbane. Al igual que en Canberra, las restricciones impuestas en Brisbane no implicaron un confinamiento, pero un porcentaje importante de la población escogió teletrabajar. El resultado fue, al igual que en Canberra, una reducción de un 13% de la concentración de NO₂ (p. 15). El periodo sucesivo, denominado fase 1, en el que se eliminaron todas las restricciones paulatinamente, se observa un aumento de las concentraciones de un 10,18% en Monash y un 7,42% en Florey. Ello es coherente con una eliminación progresiva de las restricciones puesto que los niveles de NO₂ aumentan respecto al periodo de restricciones, debido a la vuelta a la actividad industrial normal.

De la misma forma, se ha identificado que los fines de semana, al reducirse el tráfico vehicular y haber una menor actividad industrial y comercial (Escudero, 2021, p.18), la concentración de NO₂ disminuye un promedio de un 4,48% los sábados y un 5,95 % los domingos.

Finalmente, los resultados indican que, durante los periodos de incendios, se produjo una disminución del NO₂ en un 2,78% y 3,93% en Monash y Florey respectivamente. Aunque este resultado pueda parecer contrario a la regla, un estudio del Instituto de Tecnología de Massachusetts ha demostrado que durante los grandes incendios de Australia en el verano de 2019-2020, las concentraciones de NO₂ se redujeron debido a la cantidad de humo que se emitió durante los incendios (Solomon et al., 2022, p. 1).

Respecto a la partícula O₃, es un gas altamente reactivo y un contaminante secundario que se forma en la atmósfera a partir de reacciones químicas entre los óxidos de nitrógeno (NO_x) y los compuestos orgánicos volátiles (COV) en presencia de la luz solar. Estas reacciones fotoquímicas ocurren principalmente en condiciones de alta radiación solar y altas temperaturas. Los contaminantes NO_x y COV que forman el ozono se liberan en la atmósfera principalmente debido a las actividades humanas, como la combustión de combustibles fósiles, las emisiones industriales y las emisiones de los vehículos. (EEA, 2016)

Los resultados obtenidos en las tres estaciones indican que a medida que la temperatura aumenta, las concentraciones de O₃ también aumentan en un promedio de un 1,68%, lo cual se debe a que las altas temperaturas aceleran las reacciones de oxidación de los precursores de O₃, como los óxidos de nitrógeno (NO_x) y los compuestos orgánicos volátiles (COV) (Benavides, 2010). En este sentido, el aumento de las precipitaciones también aumenta la concentración de O₃ en un promedio de 0,6%. Yoo et al., (2014) explican que el O₃ es el contaminante menos afectado por las precipitaciones y que incluso, este contaminante aumenta su concentración ante las lluvias debido a “la mezcla vertical asociada con la convección que conduce al transporte descendente de O₃ desde la estratosfera inferior/troposfera superior” (p. 236).

Respecto a los periodos de restricciones impuestos durante el COVID-19, se produce un aumento del O₃ de 14,98% en Monash, 18,81% en Florey y 3,85% en Civic durante el confinamiento de Canberra entre los meses de agosto a octubre del 2021. Diversos estudios (Campbell et al., 2021; Tavella y Rodrigues, 2021; Zhang et al., 2020) analizan el aumento de la concentración de O₃ en la atmósfera durante el periodo de confinamiento frente a la reducción de los demás contaminantes atmosféricos en

diferentes ciudades. La conclusión a la que llegan estos estudios es que a pesar de que las emisiones de los contaminantes precursores del O₃ (NO_x y COV) se redujeron, se dieron unas circunstancias climatológicas favorables que promovieron la formación y concentración de O₃. Emmerson y Keywood (2021) encontraron en su estudio que la ciudad de Victoria también sufrió un aumento de O₃ durante su confinamiento frente a la reducción de los demás contaminantes atmosféricos (p. 17).

Así mismo, durante los sábados, la concentración de O₃ aumento en Monash, Florey y Civic un 5,11%, un 4,52% y un 1,36% respectivamente. Esto se debe al “Efecto fin de semana”. Mientras que la reducción de emisiones conduce a una reducción de la concentración de NO₂ en el ambiente, el O₃ aumenta. Durante los incendios, la concentración de O₃ también aumentó un 6,07% en Florey y un 0,93% en Civic. Los incendios liberan grandes cantidades de COV a la atmósfera, que actúan como precursores de O₃ y contribuyen a su formación (Agencia Estadounidense del Medio Ambiente, 2022).

En cuanto a las PM₁₀, son partículas sólidas o líquidas suspendidas en el aire que tienen un diámetro igual o menor a 10 micrómetros. Estas partículas son consideradas como partículas gruesas y pueden incluir polvo, hollín, cenizas, partículas metálicas y otros materiales. Las PM₁₀ son un tipo de contaminante atmosférico que puede ser emitido tanto por fuentes naturales como por actividades humanas. Entre las fuentes naturales se encuentran la erosión del suelo, las tormentas de polvo, los incendios forestales y las emisiones volcánicas. Por otro lado, las actividades humanas como la quema de combustibles fósiles, la emisión de gases y partículas industriales, la construcción, el tráfico vehicular y la quema de biomasa también son importantes fuentes de emisión de PM₁₀. (Gobierno del Estado de California, s.f.)

Se ha identificado que la concentración de PM₁₀ en el aire aumenta un promedio de un 3,37% cuando por cada grado centígrado que aumenta la temperatura, pero no se ha encontrado ningún informe que relacione el aumento de temperatura con el aumento de PM₁₀. De hecho, en el estudio de Acosta et al., (2019) se demuestran resultados contrarios a los obtenidos en este estudio, en dicho estudio tratan de afirmar la correlación entre la temperatura y la concentración de PM₁₀. En relación con las precipitaciones, se observa que tienen un efecto inverso en las concentraciones de PM₁₀. Por cada milímetro

de lluvia, se produce una disminución en las concentraciones de PM₁₀, reduciéndose un promedio de un 2,19 %. Según Hayas (1991), esto se debe a que la lluvia actúa como un mecanismo de limpieza del aire y arrastra las partículas al suelo, lo que reduce su presencia en la atmósfera (p. 197).

Durante los periodos de grandes incendios forestales del verano de 2019-2020, se observó un aumento significativo de un promedio de un 22,5% en las concentraciones de PM₁₀ en todas las estaciones. Esto es comprensible ya que los incendios liberan una gran cantidad de partículas finas al aire. Esta información contrasta los resultados obtenidos por otros estudios como el de Emmerson y Keywood (2021) en el que se establece que, entre diciembre de 2019 y enero de 2020, Canberra experimentó durante 45 días el peor aire exterior en comparación con cualquier otra parte del mundo (p. 57). En contraste, el informe sobre la calidad del aire realizado por IQAir sitúa en 2021 a Canberra como una de las ciudades con mejor calidad del aire (p. 10).

En cuanto al COVID-19, en los resultados se puede observar que durante la fase 1, es decir, durante el periodo en el que las restricciones fueron eliminadas, se produce un aumento promedio significativo de 25,31% en la concentración de PM₁₀. Esto podría deberse a la reactivación de las actividades humanas, incluyendo el aumento del tráfico vehicular y la actividad industrial, lo que conlleva a un aumento en las emisiones de contaminantes y, por ende, a un aumento en las concentraciones de PM₁₀. Los datos obtenidos para la fase 0, es decir, para el periodo en el que se imponen las restricciones indican que se produjo un aumento promedio de la concentración de PM₁₀ de un 7,2%. Es importante tener en cuenta que este periodo se inicia el 16 de marzo de 2020, tras los grandes incendios de diciembre 2019- enero 2020, por tanto, los niveles elevados de PM₁₀ durante la fase 0 están asociados a dichos incendios. Así lo explica Graham et al. (2021) en su estudio del impacto de los incendios de Australia.

CAPÍTULO V: CONCLUSIONES

El presente trabajo analiza los niveles de contaminación en la capital del país australiano Canberra. El trabajo se ha acotado al sexenio que abarcan los años 2017-2022. Para el estudio de la concentración de contaminación se han usado los datos obtenidos

durante el periodo seleccionado por tres estaciones de medición de contaminación localizadas en Canberra (Monash, Florey y Civic). El periodo estudiado incluye los años marcados por la pandemia del COVID-19 por lo que también se han obtenido resultados relativos a la influencia de las restricciones impuestas a raíz de la pandemia en la concentración de los contaminantes atmosféricos. Asimismo, también tienen gran influencia en los resultados la época de incendios característica de Australia y particularmente severa en los últimos años.

Una de las principales contribuciones de este estudio es la constatación del efecto de los incendios en las concentraciones de múltiples contaminantes. Canberra, y en general Australia, está condicionada por los incendios anuales destacando los grandes incendios sin precedentes del verano de 2019 – 2020. Estos grandes incendios quemaron más 10 millones de hectáreas y provocaron pérdidas catastróficas de vidas humanas, vegetación y especies (Graham et al., 2021). En cuanto a la contaminación, los niveles de concentración de PM_{10} aumentaron en promedio un 22,55% durante el periodo de estudio de este trabajo. Los altos niveles de PM_{10} son una gran amenaza para la salud humana al estar directamente relacionado con reacciones del sistema inmunitario, irritación pulmonar, células dañadas, asma, hospitalizaciones debido a enfermedades respiratorias y cardiovasculares, bronquitis crónica y cáncer (Omidi et al., 2017). En concreto, los grandes incendios de Australia provocaron 171 muertes octubre debido a la exposición a corto plazo al PM_{10} (Graham et al., 2021). Además, se ha constatado que los niveles de O_3 aumentan ante los incendios ya que estos producen emisiones de sus precursores que reaccionan en presencia de la luz solar para formar este contaminante secundario.

El estudio de los efectos de las restricciones impuestas a raíz del COVID-19 también arroja luz sobre la interacción que se da entre las distintas partículas contaminantes y la relación que se da con sus fuentes de emisión. La imposición de restricciones supuso la reducción de la emisión producida por actividades industriales y el tráfico vehicular. Aunque los resultados de esta reducción fueron positivos para la concentración de NO_2 que llegó a reducirse durante el confinamiento de Canberra en un promedio del 7,19%, las concentraciones de O_3 aumentaron en ese mismo periodo un 12,55%. A pesar de que la reducción de sus precursores debería fomentar la reducción del O_3 , otras variables como el clima tienen una gran influencia en su concentración. Además, el ozono es un contaminante secundario, lo que significa que no se emite

directamente desde una fuente específica, sino que se forma a partir de reacciones químicas complejas entre los precursores y otros compuestos presentes en la atmósfera. Esto hace que la relación entre las emisiones de contaminantes y las concentraciones de ozono sea más difícil de predecir y controlar.

Finalmente, los resultados obtenidos por el estudio también reflejan la gran relevancia que factores como la temperatura y las precipitaciones pueden tener en las concentraciones de contaminantes en la atmósfera. No todos los contaminantes se comportan de la misma forma ante un aumento de temperaturas y la presencia de precipitaciones. Mientras que el aumento de las temperaturas favorece a la reducción del NO₂ en la atmósfera, la concentración de O₃ se ve favorecida significativamente por su aumento. En concreto, en Canberra, el aumento de la temperatura de un grado centígrado supuso durante el sexenio estudiado el aumento de la concentración O₃ en un promedio de 1,68%. Asimismo, las precipitaciones redujeron en Canberra la concentración de NO₂ atmosférico en un promedio de 0,73% por milímetro de lluvia y de PM₁₀ en un 2,19%. Sin embargo, el ozono aumentó un 0,6% su concentración por cada milímetro de lluvia. Resulta relevante destacar que la lluvia no elimina la contaminación, sino que limpia la concentración atmosférica de contaminantes al empujar estos hacia el suelo (Hayas, 1991).

Como limitaciones al trabajo, debe destacarse que no se ha realizado un estudio de todas aquellas variables que pueden influir en la calidad del aire y que pueden repercutir en las concentraciones de los contaminantes en la atmósfera. Es fundamental reconocer que cualquier estudio de los efectos de las restricciones impuestas en la calidad del aire debe considerar la amplia gama de variables que pueden influir en las concentraciones de contaminantes atmosféricos.

Como futura línea de investigación del trabajo, podría implementarse un estudio sobre la influencia del relieve de Canberra en la concentración de los contaminantes atmosféricos. La situación geográfica de Canberra promueve la mayor concentración de los contaminantes sobre su atmósfera al estar situado entre dos grandes cordilleras. Además, la valoración de la radiación ultravioleta a la que está expuesta la ciudad de Canberra puede tener gran relevancia al promover la formación y estancamiento del NO₂ entre otros contaminantes.

A raíz de lo expuesto, la conclusión es que es necesario adoptar un enfoque integral que aborde tanto las fuentes de emisión de contaminantes como los factores climáticos y las condiciones locales que influyen en su concentración para poder adoptar medidas coherentes para la lucha contra la contaminación.

BIBLIOGRAFÍA

Artículos académicos

- Acosta, M., Rodríguez, L., Cuitiva, V., Socorro, A. y Aragon, J. M. (2019) *Comparación de la calidad de aire por PM 10 y PM 2.5 con la temperatura y precipitaciones en una zona aledaña a la universidad libre sede bosque en Bogotá D.C.* Asociación Colombiana de Facultades de Ingeniería (ACOFI).
- Alahmad, B., Khraisha, H., Althalji, K., Borchert, W., Al-Mulla, F., & Koutrakis, P. (2023). Connections Between Air Pollution, Climate Change & Cardiovascular Health. *Canadian Journal of Cardiology. Canadian Journal of Cardiology.*
- Albayati, N., Waisi, B., Al-Furaiji, M., Kadhom, M. y Alalwane, H. (2021). Effect of COVID-19 on air quality and pollution in different countries. *Journal of Transport & Health.*
- Anwar, Y., Rehman, K., Alharby, H. y Alghamd, K. (2018). Environmental Contamination and Remediation. *Cambridge Scholars Publishing.*
- Aránguez, E., Ordóñez, J., Serrano, J., Aragonés, N., Fernández-Patier, R., Gandarillas, A. y Galán, I. (1999). Contaminantes Atmosféricos y su vigilancia. *Rev Esp Salud Pública*, 73. pp. 123-132.
- Balasooriya, N. N., Bandara, J. S., & Rohde, N. (2022). Air pollution and health outcomes: Evidence from Black Saturday Bushfires in Australia. *Social Science & Medicine*, 306, 115165.
- Benavides, H. (2010). *Análisis de variables en altura con base en los ozonosondeos realizados en Bogotá.* Instituto de Hidrología, Meteorología y Estudios Ambientales.
- Beniston, M. (2010). Climate change and its impacts: growing stress factors for human societies. *International Review of the Red Cross*, 92, 876, pp. 557-568.

- Berger, V y Zhou, Y (2014). *Kolmogorov–Smirnov Test*. Wiley Online Library.
- Bolaño, S., Camargo, Y., Tovar, F., y Bolaño, T. (2022). The Effect of Forest Fire Events on Air Quality: A Case Study of Northern Colombia. *Fire*, 5.
- Campbell, P., Tong, D., Tang, Y., Baker, B., Lee, P., Saylor, R., Stein, A., Ma, S., Lamsal, L., y Qu, Z. (2021). Impacts of the COVID-19 economic slowdown on ozone pollution in the U.S. *Atmospheric Environment*, 264.
- Carrollo, M. C. (2012). *Regresión lineal simple*. Universidad de Santiago de Compostela. Departamento de estadística e investigación operativa.
- Chossière, P., Xu, H., Dixit, Y., Isaacs, S., Eastham, S., Allroggen, F., Speth, R. y Barrett, S. (2021). Air pollution impacts of COVID-19–related containment measures. *Science Advances*, 8, 21.
- D’Amato, G. y Akdis, C. (2020). Global warming, climate change, air pollution and allergies. *Allergy*, 75,9, pp. 2158-2160.
- Duhanyan, N. y Roustan, Y. (2011). Below-cloud scavenging by rain of atmospheric gases and particulates. *Atmospheric Environment*, 45, 39, pp. 7201-7217.
- Escudero, M., Massagué, J. y Querol, X. (2021). *Análisis de series temporales de los contaminantes atmosféricos (NO, NO2 Y O3 en la comunidad autónoma de Aragón. Variabilidad Temporal*. Universidad de Zaragoza
- Feng, X. y Wang, S. (2012). Influence of different weather events on concentrations of particulate matter with different sizes in Lanzhou, China. *Environmental Sciences*, 24, 4, pp. 665-674.
- García, M., Ramírez, H., Ulloa, H., Pérez, A. y Arias, S. (2012). Las inversiones térmicas y la contaminación atmosférica en la Zona Metropolitana de Guadalajara (México). *Investigaciones Geográficas*, 58, pp. 9-29.

- Graham, A., Pringle, K., Pope, R., Arnold, S., Conibear, L., Burns, H., Rigby, R., Arriagada, N., Butt, E., Kiely, L., Reddington, L., Spracklen, D., Woodhouse, M., Knote, C. y McQuaid, J. (2021). Impact of the 2019/2020 Australian Megafires on Air Quality and Health. *GeoHealth*, 5, 10.
- Halunga, A., Orme, C. y Yamagata, T. (2017). A heteroskedasticity robust Breusch–Pagan test for Contemporaneous correlation in dynamic panel data models. *Journal of Economics*, 198, 2.
- Hayas, A. (1991) Meteorología y contaminación atmosférica. peculiaridades de la zona urbana de Jaén. *B.I.E.G.*, 143, pp.191-200.
- Kumar, A. (2010). *Air Quality*. IntechOpen.
- Manisalidis, I., Stavropoulou, E., Stavropoulos, A. y Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A Review. *Frontiers Public Health*, 8, 14.
- Martínez, J. (2021). *Análisis de la dinámica de los contaminantes (NO₂ - PM 2.5) en el Área Metropolitana de Madrid a partir de la monitorización de la Isla de Calor Urbana (ICU) y su relación con la COVID-19*. Universidad Complutense de Madrid.
- Molina, L, Kolb, C., Foy, B., Lamb, B., Brune, W., Jimenez, J., Ramos, R., Sarmiento, J., Paramo, V., Cardenas, B., Gutierrez, V. y Molina, M. (2007). Air quality in North America's most populous city – overview of the MCMA-2003 campaign. *Atmospheric. Chemistry. Physics* 7.
- Omidi, Y., Jundishapur, A. y Borgini, A. (2017). Exposure to PM₁₀, NO₂ and O₃ and impacts on human health. *Springer*.
- Pabón, J.D., Zea, J., León, G, Hurtado, G. y Edgar, J. (s.f.). *La atmosfera, el tiempo y el clima*. IDEAM.

- Rahman, S., Chang, H., Magill, C., Tomkins, K. (2018). *Forest Fire Occurrence and Modeling in Southeastern Australia*. Intech
- Rojo, J. M. (2007). *Regresión lineal múltiple*. Instituto de Economía y Geografía, Laboratorio de Estadística, Editor, 32.
- Saha, L., Kumar, A., Kumar, S., Korstad, J., Srivastava, S. y Baudh, K. (2022) The impact of the COVID-19 lockdown on global air quality: A review. *Environmental Sustainability*, pp. 5-23.
- Sario, M., Katsouyanni, K. y Michelozzi, P. (2013). Climate change, extreme weather events, air pollution and respiratory health in Europe. *European Respiratory Journal*, 42, pp. 826-843.
- Schraufnagel, D., Balmes, J., Cowl, C., De Matteis, S., Jung, S., Mortimer, K., Perez, R., Rice, M., Riojas, H., Sood, A., Thurston, G., To T, Vanker, A., Wuebbles, D. (2019). Air Pollution and Noncommunicable Diseases. *Chest*, 155, 2, pp. 417-426.
- Shahzad, U. (2015). Global Warming: Causes, Effects and Solutions. *Durreesamin Journal*, 4, 1.
- Sharma, D. y Mauzerall, D. (2022). Analysis of Air Quality Data in India between 2015 and 2019. *Aerosol and Air Quality Research*, 22.
- Seguel, R., Morales, R., Leiva, M. (2012). Ozone weekend effect in Santiago, Chile. *Environmental pollution*, 162, pp. 72-79.
- Sicard, P., Paoletti, E., Agathokleous, E., Araminienè, V., Proietti, C., Coulibaly, F. y De Marco, A. (2020). Ozone weekend effect in cities: Deep insights for urban air pollution control. *Environmental Research*, 191.
- Solomon, S., Dube, K., Stone, K. Yu, P., Kinnison, D., Toon, O., Strahan, S., Rosenlof, K., Portmann, R., Davis, S., Randel, W., Bernath, P, Booneh, C., Bardeen, C.,

Bourassa, A., Zawada, D. y Degenstein, D. (2022). On the stratospheric chemistry of midlatitude wildfire smoke. *PNAS*, 119, 10, pp. 1-9.

Tavella, R. y Rodrigues, M. (2021). Watch out for trends: did ozone increased or decreased during the COVID-19 pandemic? *Environ Sci Pollut Res Int*, 28.

Tilt, B. (2019). China's air pollution crisis: Science and policy perspectives. *Environmental Science and Policy*, 92.

Vasconcellos, L., Adler, R., Da Silva, A., Santos, J. y Rodrigues, F. (2022). Weekend effect on air pollutant levels in southernmost cities of Brazil with different economic activities. *Environmental monitoring and assessment*, 194.

Williamson, R., Banwell, C., Calear, A. y LaBond, C. (2022). Bushfire Smoke in Our Eyes: Community Perceptions and Responses to an Intense Smoke Event in Canberra, Australia. *Frontiers in Public Health*.

Informes

Agencia Estadounidense del Medio Ambiente (2022). *Wildland Fire Activity and Modeled Impacts on O3 and PM2.5*. Air Quality Assessment Division. EPA -454/R -22 -00 2.

Cancer Council (s.f.). *UV Index*. Disponible en: <https://www.cancer.org.au/cancer-information/causes-and-prevention/sun-safety/uv-index>

Emmerson, M. y Keywood, D. (2021). *Australia state of the environment 2021: air quality*. Independent report to the Australian Government Minister for the Environment, Commonwealth of Australia, Canberra. Disponible en: <https://soe.dcceew.gov.au/sites/default/files/2022-07/soe2021-air-quality.pdf>

Gobierno del Territorio de la Capital Australiana (s.f.). *Canberra—All you need to know about the nation's capital*. Disponible en:

https://www.theclimategroup.org/sites/default/files/2020-10/Australian-Capital-Territory-Appendix_0.pdf

Grupo Intergubernamental de Expertos sobre el Cambio Climático (2021). *Climate Change 2021: The physical Science Basis*. Working group. Disponible en: https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM_final.pdf

IQAir (2021). *World Air Quality Report*. Disponible en: [file:///C:/Users/carlo/Downloads/world-air-quality-report-2021-en%20\(1\).pdf](file:///C:/Users/carlo/Downloads/world-air-quality-report-2021-en%20(1).pdf)

Junta de Recursos del Aire de California (2021). *Camp Fire Air Quality Data Analysis*. División de Ciencia y Planificación de la Calidad del Aire. Disponible en: https://ww2.arb.ca.gov/sites/default/files/2021-07/Camp_Fire_report_July2021.pdf

Kenward, A, Adams-Smith, D. y Raja, U. (2013). *Wildfires and Air Pollution*. Climate Central. Disponible en: <http://assets.climatecentral.org/pdfs/WildfiresAndAirPollution.pdf>

Organización Mundial de la Salud (2022). *Compendium of WHO and other UN guidance on health and environment*. Department of Environment, Climate Change and Health. Disponible en: https://cdn.who.int/media/docs/default-source/environmental-health-impact/who_compendium_air_pollution_01042022_eo_final.pdf?sfvrsn=1c4ceb5_3

Vassarotti, R. (2021). *Bushfire Smoke and Air Quality Strategy 2021-2025*. Minister for the Environment of ACT Government. Disponible en: https://www.parliament.act.gov.au/__data/assets/pdf_file/0018/1900323/Bushfire-Smoke-and-Air-Quality-Strategy-2021-2025.pdf

- Wang, R., Cui, K., Sheu, H., Wang, L. y Liu¹, X. (2023). Effects of Precipitation on the Air Quality Index, PM_{2.5} Levels and on the Dry Deposition of PCDD/Fs in the Ambient Air. *Aerrol and Air Quality Research*, 24, 4.
- Yadav, D., Rangabhashiyam, S., Verma, P., Singh, P., Devi, P., Kumar, P., Hussain, C., Gaurav, G., Kumar, K. (2021). Environmental and health impacts of contaminants of emerging concerns: Recent treatment challenges and approaches. *Chemosphere*, 272.
- Yoo, J., Lee, Y., Kim, D., Jeong, M., Stockwell · w., Kundu, K., Oh, S., Shin, D. y Lee, S. (2014). New indices for wet scavenging of air pollutants (O₃, CO, NO₂, SO₂, and PM₁₀) by summertime rain. *Atmospheric Environment*, 82, pp. 226-237.
- Zhang, K., Liu, Z., Zhang, X., Li, Q., Jensen, A., Tan, W., Huang, L., Wang, Y., Gouw, J. y Li. L. (2022). *Insights into the significant increase in ozone during COVID-19 in a typical urban city of China. European Geosciences Union*, 22, 7

Recursos de internet

- Agencia Europea del Medio Ambiente. (19 de abril de 2016). *Ozono troposférico*. El Medio Ambiente en Europa: Segunda evaluación. Disponible en: <https://www.eea.europa.eu/es/publications/92-828-3351-8/page005.html#:~:text=El%20ozono%20se%20forma%20en,de%20NOx%20y%20luz%20solar>
- Agencia Europea del Medio Ambiente (2020). *Contaminación atmosférica*. Disponible en: <https://www.eea.europa.eu/downloads/614d5dfa214c977868c097af826fc6f0/1606129146/intro.pdf>
- Agencia Europea del Medio Ambiente (5 de mayo de 2023). *How air pollution affects our health*. Disponible en: <https://www.eea.europa.eu/en/topics/in-depth/air-pollution/eow-it-affects-our-health>

Agencia Estadounidense del Medio Ambiente (2023). *Sources of Greenhouse Gas Emissions*. Greenhouse gas emissions. Disponible en: [https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions#:~:text=Industry%20\(23%25%20of%202021%20greenhouse,produce%20goods%20from%20raw%20materials](https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions#:~:text=Industry%20(23%25%20of%202021%20greenhouse,produce%20goods%20from%20raw%20materials)

Agencia Estadounidense del Medio Ambiente (20 de marzo de 2023). *Nitrogen Dioxide's Impact on Indoor Air Quality*. Disponible en: <https://www.epa.gov/indoor-air-quality-iaq/nitrogen-dioxides-impact-indoor-air-quality#:~:text=Health%20Effects%20Associated%20with%20Nitrogen%20Dioxide,-EPA's%20Integrated%20Risk&text=NO2%20acts%20mainly%20as,edema%20and%20diffuse%20lung%20injury>

Agencia Estadounidense del Medio Ambiente (30 de agosto de 2022). *Health and Environmental Effects of Particulate Matter (PM)*. Disponible en: <https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm>

Agencia Estadounidense del Medio Ambiente (24 de mayo de 2023). *Health Effects of Ozone Pollution*. Disponible en: <https://www.epa.gov/ground-level-ozone-pollution/health-effects-ozone-pollution>

Bodart, B. (2022). *Walter Burley Griffin's Canberra and his Tramway. The Ideal City of the Future?* Canberra Planning Action Group. Disponible en: <https://www.canberraplanningactiongroup.com/post/walter-burley-griffin-s-canberra-and-his-tramway-the-ideal-city-of-the-future-part-one-of-three>

Bureau of Meteorology (s.f.). *Australian Capital Territory Weather and Warnings*. Disponible en: <http://www.bom.gov.au/act/>

Datademia (2022). *¿Qué es R?* Disponible en: <https://datademia.es/blog/que-es-r>

Departamento de Salud y Control Ambiental de Carolina del Sur (s.f.). *How is Ozone Formed?* Disponible en: <https://scdhec.gov/environment/your-air/most-common-air-pollutants/about-ozone/how-ozone-formed#:~:text=Ozone%20is%20a%20molecule%20made,are%20also%20known%20as%20Hydrocarbons>

Ember (2022). *Per Capita Coal Power Emissions, 2022*. Disponible en: <https://ember-climate.org/insights/research/per-capita-coal-power-emissions-2022/>

Gobierno Australiano (29 de Julio de 2022). *National Greenhouse and Energy Reporting Scheme*. Department of Climate Change, Energy, the Environment and Water. Disponible en: <https://www.dcceew.gov.au/environment/protection/chemicals-management/international-agreements/stockholm-convention>

Gobierno Australiano (22 de agosto de 2022). *National Clean Air Agreement*. Department of Climate Change, Energy, the Environment and Water. Disponible en: <https://www.dcceew.gov.au/environment/protection/air-quality/national-clean-air-agreement>

Gobierno Australiano (27 de febrero de 2023). *Stockholm Convention on Persistent Organic Pollutants (POPs)*. Department of Climate Change, Energy, the Environment and Water. Disponible en: <https://www.dcceew.gov.au/environment/protection/chemicals-management/international-agreements/stockholm-convention>

Gobierno Australiano (5 de mayo de 2023). *Safeguard Mechanism*. Department of Climate Change, Energy, the Environment and Water. Disponible en: <https://www.dcceew.gov.au/climate-change/emissions-reporting/national-greenhouse-energy-reporting-scheme/safeguard-mechanism>

Gobierno del Estado de California (s.f.). *Inhalable Particulate Matter and Health (PM2.5 and PM10)*. California Air Resources Board. Disponible en: <https://ww2.arb.ca.gov/resources/inhalable-particulate-matter-and->

Universidad Internacional de Valencia (4 de octubre de 2022). *Lenguaje de programación R: qué es, características e importancia en el Big Data*. Disponible en: <https://www.universidadviu.com/es/actualidad/nuestros-expertos/lenguaje-de-programacion-r-que-es-caracteristicas-e-importancia-en-el-big-data>

Universidad de Santiago de Compostela (2011). *Regresión Lineal Múltiple El modelo, estimación de los parámetros, contrastes*. Disponible en: http://eio.usc.es/eipc1/BASE/BASEMASTER/FORMULARIOS-PHP-DPTO/MATERIALES/Mat_50140128_RegresionMultiple.pdf

Walter. C (2019). *The public health impacts of air pollution in Australia: Research, policy and planning disconnects*. Universidad de Melbourne. Disponible en: <https://www.climatecollege.unimelb.edu.au/seminar/public-health-impacts-air-pollution-australia-research-policy-and-planning-disconnects>

ANEXOS

```
library(tidyverse) # para tratar los datos,dataframes y mas
library(lubridate) # para tratar la variable tiempo

#Para la carga correcta de los datos guarde el script y los datos en la misma carpeta
setwd(dirname(rstudioapi::getActiveDocumentContext())$path)
getwd()

data <- read_csv("Air_Quality_Monitoring_Data.csv",
  col_types = cols(NO2 = col_double(),
    O3_1hr = col_double(), AQI_CO = col_integer(),
    AQI_NO2 = col_integer(), AQI_O3_1hr = col_integer(),
    AQI_O3_4hr = col_integer(), AQI_O3_8hr = col_integer(),
    AQI_PM10 = col_integer(), AQI_PM2.5 = col_integer(),
    AQI_Site = col_integer(), Time = col_character()))

# Evauacion de la Calidad - tratamiento de los datos y valores faltantes

str(data) # Tipo de variable

# Creando variables de Tiempo

data$DateT <- as.Date(data$DateTime, format = " %d/%m/%Y %T ")
data$DateTimeHour <- parse_date_time(data$DateTime, "%d/%m/%y %I:%M:%S
%p") # crea una nueva variable con AÑO/MES/DIA hr:min:seg
data$Anio <- as.numeric(format(data$DateT,'%Y'))

# Filtramos el periodo que nos interesa 2017-2022
data <- data %>% filter(DateT>= "2017-01-01" & DateT <="2022-12-31")
```

```

# Filtramos las variables que nos interesan para el analisis: PM10, NO2, O3_1hr,
AQI_O31hr, AQIPM10, AQINO2, AQI SITE
data <- data%>% select
(Name,'GPS','DateT','DateTimeHour','Anio','NO2','O3_1hr','PM10','AQI_NO2','AQI_
O3_1hr','AQI_PM10','AQI_Site')

# Grafico de la distribucion de los datos en general por particula
# BoxPlot
library(ggplot2)
par(mfrow=c(2,3))
hist(x = data$NO2,main="Histograma de NO2" , xlab = "NO2", ylab =
"Frecuencia",col = "purple")
hist(x = data$O3_1hr, main = "Histograma de O3", xlab = "O3", ylab =
"Frecuencia",col = "purple")
hist(x = data$PM10, main = "Histograma de PM10", xlab = "PM10", ylab =
"Frecuencia",col = "purple")
boxplot(data$NO2, col = "green")
boxplot(data$O3_1hr, col = "green")
boxplot(data$PM10, col = "green")

#funcion que agrupa los datos por estación
estacion<-function(estacion){
  df = data[data$Name == estacion, ]
  assign(x = estacion, value = df, envir = globalenv())
}

estacion("Monash") #crea dataframe de Monash con todas las variables
estacion("Florey") #crea dataframe de Florey con todas las variables
estacion("Civic") #crea dataframe de Civic con todas las variables

#-----
# Monash

```

```

#-----
# Imputacion de los valores faltantes
#-----

library(mice) # permite realizar la imputacion de los datos mediante diferentes técnicas
library(VIM) # permite visualizar los datos faltantes
library(naniar) # Muestra el total de valores faltantes

# Metodo con funcion naniar
miss_var_summary(Monash)

# El metodo norm.boot, realiza la imputacion con regresiones y de esa manera fortalece
por simulacion la relacion en los datos
# Se procedera a imputar agrupando variables, por conveniencia, segun su escala
imputed_data_0 <- mice(Monash%>%select('NO2', 'O3_1hr', 'PM10'), method =
"norm")
imputed_data_1 <- mice(Monash%>%select('AQI_NO2', 'AQI_O3_1hr', 'AQI_PM10',
'AQI_Site'), method = "norm.boot")

# Se asigna a una variable cada imputacion
Data_Impu_0 <- mice::complete(imputed_data_0)
Data_Impu_1 <- mice::complete(imputed_data_1)

#Color rojo es el valor verdadero y verde es la imputacion
par(mfrow=c(1,2))
plot(density(Monash$NO2,na.rm = T),col=2,main="NO2")
lines(density(as.numeric(Data_Impu_0$NO2)),col=3)
plot(density(Monash$O3_1hr,na.rm = T),col=2,main="O3_1hr")
lines(density(as.numeric(Data_Impu_0$O3_1hr)),col=3)

par(mfrow=c(1,2))
plot(density(Monash$PM10,na.rm = T),col=2,main="PM10")
lines(density(as.numeric(Data_Impu_0$PM10)),col=3)
plot(density(Monash$AQI_NO2,na.rm = T),col=2,main="AQI_NO2")

```

```

lines(density(as.numeric(Data_Impu_1$AQI_NO2)),col=3)

par(mfrow=c(1,2))
plot(density(Monash$AQI_O3_1hr,na.rm = T),col=2,main="AQI_O3_1hr")
lines(density(as.numeric(Data_Impu_1$AQI_O3_1hr)),col=3)
plot(density(Monash$AQI_PM10,na.rm = T),col=2,main="AQI_PM10")
lines(density(as.numeric(Data_Impu_1$AQI_PM10)),col=3)

plot(density(Monash$AQI_Site,na.rm = T),col=2,main="AQI_Site")
lines(density(as.numeric(Data_Impu_1$AQI_Site)),col=3)

# Se reemplaza las variables imputadas en la base original.
Monash<-Monash%>%mutate(NO2 = Data_Impu_0$NO2, O3_1hr =
Data_Impu_0$O3_1hr, PM10 = Data_Impu_0$PM10
, AQI_NO2 = Data_Impu_1$AQI_NO2, AQI_O3_1hr =
Data_Impu_1$AQI_O3_1hr
, AQI_PM10 = Data_Impu_1$AQI_PM10, AQI_Site =
Data_Impu_1$AQI_Site)

# Se comprueba que no quedan valores faltantes
miss_var_summary(Monash)

# Extraemos los datos para cada particula NO2, O3 y PM10 de Monash
# Agrupamos los datos por promedio de particula de dia, mes y año de los periodos
2017-2022
Monash_NO2 <- Monash %>% mutate(year = format(DateT, "%Y"), month =
format(DateT, "%m"), day = format(DateT, "%d")) %>%
group_by(year, month, day) %>%
summarize(mean_NO2 = mean(NO2))

Monash_O3 <- Monash %>% mutate(year = format(DateT, "%Y"), month =
format(DateT, "%m"), day = format(DateT, "%d")) %>%
group_by(year, month, day) %>%

```

```

summarize(mean_O3 = mean(O3_1hr))

Monash_PM10 <- Monash %>% mutate(year = format(DateT, "%Y"), month =
format(DateT, "%m"), day = format(DateT, "%d")) %>%
  group_by(year, month, day) %>%
  summarize(mean_PM10 = mean(PM10))

# GRAFICO PARA VER EL COMPORTAMIENTO DIARIO POR MES
# 2017-2018
avg_tb_NO2_0 <- Monash_NO2 %>% group_by(month, day) %>% filter(year>=
"2017" & year <="2018") %>%
  summarize(S2017_2018 = mean(mean_NO2))
avg_tb_NO2_0 <- avg_tb_NO2_0 %>% unite(day_month, day, month, sep = "/",
remove = TRUE)

avg_tb_O3_0 <- Monash_O3 %>% group_by(month, day) %>% filter(year>= "2017"
& year <="2018") %>%
  summarize(S2017_2018 = mean(mean_O3))
avg_tb_O3_0 <- avg_tb_O3_0 %>% unite(day_month, day, month, sep = "/", remove =
TRUE)

avg_tb_PM10_0 <- Monash_PM10 %>% group_by(month, day) %>% filter(year>=
"2017" & year <="2018") %>%
  summarize(S2017_2018 = mean(mean_PM10))
avg_tb_PM10_0 <- avg_tb_PM10_0 %>% unite(day_month, day, month, sep = "/",
remove = TRUE)

# 2019-2020
avg_tb_NO2_1 <- Monash_NO2 %>% group_by(month, day) %>% filter(year>=
"2019" & year <="2020") %>%
  summarize(S2019_2020 = mean(mean_NO2))
avg_tb_NO2_1 <- avg_tb_NO2_1 %>% unite(day_month, day, month, sep = "/",
remove = TRUE)

```

```

avg_tb_O3_1 <- Monash_O3 %>% group_by(month, day) %>% filter(year>= "2019"
& year <="2020") %>%
  summarize(S2019_2020 = mean(mean_O3))
avg_tb_O3_1 <- avg_tb_O3_1 %>% unite(day_month, day, month, sep = "/", remove =
TRUE)

avg_tb_PM10_1 <- Monash_PM10 %>% group_by(month, day) %>% filter(year>=
"2019" & year <="2020") %>%
  summarize(S2019_2020 = mean(mean_PM10))
avg_tb_PM10_1 <- avg_tb_PM10_1 %>% unite(day_month, day, month, sep = "/",
remove = TRUE)

# 2021-2022
avg_tb_NO2_2 <- Monash_NO2 %>% group_by(month, day) %>% filter(year>=
"2021" & year <="2022") %>%
  summarize(S2021_2022 = mean(mean_NO2))
avg_tb_NO2_2 <- avg_tb_NO2_2 %>% unite(day_month, day, month, sep = "/",
remove = TRUE)

avg_tb_O3_2 <- Monash_O3 %>% group_by(month, day) %>% filter(year>= "2021"
& year <="2022") %>%
  summarize(S2021_2022 = mean(mean_O3))
avg_tb_O3_2 <- avg_tb_O3_2 %>% unite(day_month, day, month, sep = "/", remove =
TRUE)

avg_tb_PM10_2 <- Monash_PM10 %>% group_by(month, day) %>% filter(year>=
"2021" & year <="2022") %>%
  summarize(S2021_2022 = mean(mean_PM10))
avg_tb_PM10_2 <- avg_tb_PM10_2 %>% unite(day_month, day, month, sep = "/",
remove = TRUE)

# Unimos los datos de 2020-2022 hacia 2016-2019

```

```

avg_tb_NO2 <- left_join(avg_tb_NO2_0, avg_tb_NO2_1, by ='day_month') %>%
left_join(., avg_tb_NO2_2, by='day_month')
avg_tb_O3 <- left_join(avg_tb_O3_0, avg_tb_O3_1, by ='day_month') %>% left_join(.,
avg_tb_O3_2, by='day_month')
avg_tb_PM10 <- left_join(avg_tb_PM10_0, avg_tb_PM10_1, by ='day_month') %>%
left_join(., avg_tb_PM10_2, by='day_month')

# Convierte de tibble a DataFrame
avg_NO2 <- avg_tb_NO2 %>% as.data.frame()
avg_O3 <- avg_tb_O3 %>% as.data.frame()
avg_PM10 <- avg_tb_PM10 %>% as.data.frame()

# Es necesario convertir a date el tipo de dato day_month a tipo date para graficas
avg_NO2$day_month <- as.Date(avg_NO2$day_month, format="%d/%m")
avg_O3$day_month <- as.Date(avg_O3$day_month, format="%d/%m")
avg_PM10$day_month <- as.Date(avg_PM10$day_month, format="%d/%m")

# Grafico de serie de NO2 de Monash
p= ggplot()+
  geom_line(data = avg_NO2, aes(x=day_month,
y=S2017_2018,colour="S2017_2018"))+
  geom_line(data = avg_NO2,
aes(x=day_month,y=S2019_2020,colour="S2019_2020"))+
  geom_line(data = avg_NO2,
aes(x=day_month,y=S2021_2022,colour="S2021_2022"))+
  scale_x_date(date_labels = "%B", date_breaks = 'month') +
  labs(title="Comparativa de la evolución de la media diaria de NO2 de la estacion
Monash, periodo [2017-2018], [2019-2020], [2021-2022]") +
  ylab("Media NO2") +
  xlab("Meses") +
  scale_color_manual(name = "Serie",
values = c("S2017_2018" = '#018576', "S2019_2020"='blue2',
"S2021_2022"='#CA0020', size=12))

```

```
print(p+ theme(axis.text = element_text(size=11), legend.title = element_text (size = 12), legend.text = element_text (size = 12)))
```

```
# Grafico de serie de O3 de Monash
```

```
p= ggplot()+  
  geom_line(data = avg_O3, aes(x=day_month,  
y=S2017_2018,colour="S2017_2018"))+  
  geom_line(data = avg_O3,  
aes(x=day_month,y=S2019_2020,colour="S2019_2020"))+  
  geom_line(data = avg_O3,  
aes(x=day_month,y=S2021_2022,colour="S2021_2022"))+  
  scale_x_date(date_labels = "%B", date_breaks = 'month') +  
  labs(title="Comparativa de la evolución de la media diaria de O3 de la estacion  
Monash, periodo [2017-2018], [2019-2020], [2021-2022]") +  
  ylab("Media O3") +  
  xlab("Meses") +  
  scale_color_manual(name = "Serie",  
values = c("S2017_2018" = '#018576', "S2019_2020"='blue2',  
"S2021_2022"='#CA0020'))
```

```
print(p+ theme(axis.text = element_text(size=11), legend.title = element_text (size = 12), legend.text = element_text (size = 12)))
```

```
# Grafico de serie de PM10 de Monash
```

```
p= ggplot()+  
  geom_line(data = avg_PM10, aes(x=day_month,  
y=S2017_2018,colour="S2017_2018"))+  
  geom_line(data = avg_PM10,  
aes(x=day_month,y=S2019_2020,colour="S2019_2020"))+  
  geom_line(data = avg_PM10,  
aes(x=day_month,y=S2021_2022,colour="S2021_2022"))+  
  scale_x_date(date_labels = "%B", date_breaks = 'month') +
```



```

labs(title="Comparativa de la evolución de la media diaria de PM10 de la estación
Monash, periodo [2017-2018], [2019-2020], [2021-2022]") +
ylab("Media PM10") +
xlab("Meses") +
scale_color_manual(name = "Serie",
                    values = c("S2017_2018" = '#018576', "S2019_2020"='blue2',
                                "S2021_2022"='#CA0020'))

print(p+ theme(axis.text = element_text(size=11), legend.title = element_text (size =
12), legend.text = element_text (size = 12)))

# Convierte de tibble a DataFrame
Monash_NO2 <- Monash_NO2 %>% as.data.frame()
Monash_O3 <- Monash_O3 %>% as.data.frame()
Monash_PM10 <- Monash_PM10 %>% as.data.frame()

#Exportamos a excel para unir con las otras variables dependientes
library("writexl")
write_xlsx(Monash_NO2,"Monash_NO2.xlsx")
write_xlsx(Monash_O3,"Monash_O3.xlsx")
write_xlsx(Monash_PM10,"Monash_PM10.xlsx")

summary(Monash)

# MONASH - NO2
# Leyendo la nueva data con las variables independientes agregadas
Monash_NO2_reg=read.csv("Monash_NO2.csv",header = TRUE, sep = ";")

str(Monash_NO2_reg)
Monash_NO2_reg$Fecha <- as.Date(Monash_NO2_reg$Fecha, format = "
%d/%m/%Y")

#Verificación de datos faltantes de las variables añadidas

```

```

miss_var_summary(Monash_NO2_reg)

#Imputacion de Precipitacion y temperatura
imputed_Precip_Temp <- mice(Monash_NO2_reg%>%select('Precipitacion',
'Temperatura'), method = "norm")

# Se asigna a una variable cada imputacion
imputed_Precip_Temp <- mice::complete(imputed_Precip_Temp)

par(mfrow=c(1,2))
plot(density(Monash_NO2_reg$Precipitacion,na.rm = T),col=2,main="Precipitacion")
lines(density(as.numeric(imputed_Precip_Temp$Precipitacion)),col=3)
plot(density(Monash_NO2_reg$Temperatura,na.rm = T),col=2,main="Temperatura")
lines(density(as.numeric(imputed_Precip_Temp$Temperatura)),col=3)

# Se reemplaza las variables imputadas en la base original.
Monash_NO2_reg<-Monash_NO2_reg%>%mutate(Precipitacion =
imputed_Precip_Temp$Precipitacion,
          Temperatura = imputed_Precip_Temp$Temperatura)

# Asignando el tipo de dato que le corresponde
Monash_NO2_reg$Incendio <- as.factor(Monash_NO2_reg$Incendio)
Monash_NO2_reg$Fase.0 <- as.factor(Monash_NO2_reg$Fase.0)
Monash_NO2_reg$Fase.1 <- as.factor(Monash_NO2_reg$Fase.1)
Monash_NO2_reg$Confinamiento <- as.factor(Monash_NO2_reg$Confinamiento)
Monash_NO2_reg$Desconfinamiento <-
as.factor(Monash_NO2_reg$Desconfinamiento)
Monash_NO2_reg$Sabado <- as.factor(Monash_NO2_reg$Sabado)
Monash_NO2_reg$Domingo <- as.factor(Monash_NO2_reg$Domingo)

# Visualizando las nuevas variables predictoras tipo factor
library(ggpubr) #para gestionar multiples graficos

```

```

g1 <- ggplot(data = Monash_NO2_reg, mapping=aes(x = Incendio, y = Promedio_NO2,
color=Incendio)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g2 <- ggplot(data = Monash_NO2_reg, mapping=aes(x = Fase.0, y = Promedio_NO2,
color=Fase.0)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g3 <- ggplot(data = Monash_NO2_reg, mapping=aes(x = Fase.1, y = Promedio_NO2,
color=Fase.1)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g4 <- ggplot(data = Monash_NO2_reg, mapping=aes(x = Confinamiento, y =
Promedio_NO2, color=Confinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g5 <- ggplot(data = Monash_NO2_reg, mapping=aes(x = Desconfinamiento, y =
Promedio_NO2, color=Desconfinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g6 <- ggplot(data = Monash_NO2_reg, mapping=aes(x = Sabado, y = Promedio_NO2,
color=Sabado)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g7 <- ggplot(data = Monash_NO2_reg, mapping=aes(x = Domingo, y =
Promedio_NO2, color=Domingo)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +

```

```

theme_bw() + theme(legend.position = "none")

annotate_figure(ggarrange(g1,g2,g3,g4,g5,g6,g7,
                        ncol = 4, nrow = 2),
                top=text_grob("BoxPlot de las variables tipo factor"))

# Regresion

# install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)

chart.Correlation(Monash_NO2_reg[,c('Promedio_NO2','Temperatura', 'Precipitacion')],
                  histogram = TRUE, method = "pearson")

# Modelo General
modelo <- lm(Promedio_NO2 ~ Temperatura + Precipitacion + Incendio + Fase.0 +
             Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
             Monash_NO2_reg )
summary(modelo)

# Validación de condiciones para la regresión múltiple lineal

library(ggplot2)
library(gridExtra)

# Relación lineal entre los predictores numéricos y la variable dependiente
plot1 <- ggplot(data = Monash_NO2_reg, aes(Temperatura, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
  theme_bw()
plot2 <- ggplot(data = Monash_NO2_reg, aes(Precipitacion, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
  theme_bw()

```

```

plot1 <- plot1+ theme(axis.text = element_text(size=12), axis.title = element_text(size =
16))
plot2 <- plot2+ theme(axis.text = element_text(size=12), axis.title = element_text(size =
16))

grid.arrange(plot1, plot2)

# Normalidad
qqnorm(modelo$residuals, main="Grafico QQnorm", ylab= "Cuantiles
muestrales",xlab="Cuantiles teóricos N(0,1)",cex.lab=1.3,cex.axis=1.2)
qqline(modelo$residuals)
# Test de normalidad
ks.test(modelo$residuals, "pnorm", mean = mean(modelo$residuals), sd =
sd(modelo$residuals))

# Homocedasticidad

# Variabilidad constante de los residuos:
plot1 <- ggplot(data = data.frame(predict_values = predict(modelo),
residuos = residuals(modelo)),
aes(x = predict_values, y = residuos)) +
geom_point() +
geom_smooth(color = "firebrick", se = FALSE) +
geom_hline(yintercept = 0) +
theme_bw()
plot1 <- plot1 + theme(axis.text = element_text(size=12),axis.title.x = element_text(size
= 16), axis.title.y = element_text(size = 16))+ labs(x='Valores predichos', y='Residuos')
plot1

# Test de homocedasticidad
library(lmtest)
bptest(modelo)

```

```

# Identificación de posibles valores atípicos o influyentes
library(car)
outlierTest(modelo)

summary(influence.measures(modelo))

influencePlot(modelo)

# los que ocupan las posiciones 330,336,1096,1100,1884

# Se decide rehacer el modelo excluyendo los outliers y los puntos influyentes
# que ocupan las posiciones de 1100,336,1096,330,1884,422,713,819,1097,1098,

Monash_NO2_reg <- Monash_NO2_reg[-c(1100, 1096, 336, 1097, 1098,330,1884), ]

# Modelo General
modelo <- lm(Promedio_NO2 ~ Temperatura + Precipitacion + Incendio + Fase.0 +
             Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
             Monash_NO2_reg )
summary(modelo)

# Selección de los mejores predictores
step(object = modelo, direction = "both", trace = 1)

# Mejor modelo sin puntos influyentes
modelo <- lm(log(Promedio_NO2+0.01) ~ Temperatura + Precipitacion + Incendio +
             Fase.0 + Fase.1 + Confinamiento + Desconfinamiento + Sabado +
             Domingo, data = Monash_NO2_reg)
summary(modelo)

# MONASH - O3
# Leyendo la nueva data con las variables independientes agregadas
Monash_O3_reg=read.csv("Monash_O3.csv",header = TRUE, sep = ";")

```

```

str(Monash_O3_reg)
Monash_O3_reg$Fecha <- as.Date(Monash_O3_reg$Fecha, format = " %d/%m/%Y")

#Verificacion de datos faltantes de las variables añadidas
miss_var_summary(Monash_O3_reg)

#Imputacion de Precipitacion y temperatura
imputed_Precip_Temp <- mice(Monash_O3_reg%>%select('Precipitacion',
'Temperatura'), method = "norm")

# Se asigna a una variable cada imputacion
imputed_Precip_Temp <- mice::complete(imputed_Precip_Temp)

par(mfrow=c(1,2))
plot(density(Monash_O3_reg$Precipitacion,na.rm = T),col=2,main="Precipitacion")
lines(density(as.numeric(imputed_Precip_Temp$Precipitacion)),col=3)
plot(density(Monash_O3_reg$Temperatura,na.rm = T),col=2,main="Temperatura")
lines(density(as.numeric(imputed_Precip_Temp$Temperatura)),col=3)

# Se reemplaza las variables imputadas en la base original.
Monash_O3_reg<-Monash_O3_reg%>%mutate(Precipitacion =
imputed_Precip_Temp$Precipitacion,
          Temperatura = imputed_Precip_Temp$Temperatura)

# Asignando el tipo de dato que le corresponde
Monash_O3_reg$Incendio <- as.factor(Monash_O3_reg$Incendio)
Monash_O3_reg$Fase.0 <- as.factor(Monash_O3_reg$Fase.0)
Monash_O3_reg$Fase.1 <- as.factor(Monash_O3_reg$Fase.1)
Monash_O3_reg$Confinamiento <- as.factor(Monash_O3_reg$Confinamiento)
Monash_O3_reg$Desconfinamiento <- as.factor(Monash_O3_reg$Desconfinamiento)
Monash_O3_reg$Sabado <- as.factor(Monash_O3_reg$Sabado)
Monash_O3_reg$Domingo <- as.factor(Monash_O3_reg$Domingo)

```

```

# Visualizando las nuevas variables predictoras tipo factor
library(ggpubr) #para gestionar multiples graficos
g1 <- ggplot(data = Monash_O3_reg, mapping=aes(x = Incendio, y = Promedio_O3,
color=Incendio)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g2 <- ggplot(data = Monash_O3_reg, mapping=aes(x = Fase.0, y = Promedio_O3,
color=Fase.0)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g3 <- ggplot(data = Monash_O3_reg, mapping=aes(x = Fase.1, y = Promedio_O3,
color=Fase.1)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g4 <- ggplot(data = Monash_O3_reg, mapping=aes(x = Confinamiento, y =
Promedio_O3, color=Confinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g5 <- ggplot(data = Monash_O3_reg, mapping=aes(x = Desconfinamiento, y =
Promedio_O3, color=Desconfinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g6 <- ggplot(data = Monash_O3_reg, mapping=aes(x = Sabado, y = Promedio_O3,
color=Sabado)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")

```



```

g7 <- ggplot(data = Monash_O3_reg, mapping=aes(x = Domingo, y = Promedio_O3,
color=Domingo)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")

annotate_figure(ggarrange(g1,g2,g3,g4,g5,g6,g7,
  ncol = 4, nrow = 2),
  top=text_grob("BoxPlot de las variables tipo factor"))

```

```
# Regresion
```

```
# install.packages("PerformanceAnalytics")
```

```
library(PerformanceAnalytics)
```

```
chart.Correlation(Monash_O3_reg[,c('Promedio_O3','Temperatura', 'Precipitacion')],
  histogram = TRUE, method = "pearson")
```

```
# Modelo General
```

```
modelo <- lm(log(Promedio_O3) ~ Temperatura + Precipitacion + Incendio + Fase.0 +
  Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
  Monash_O3_reg )
summary(modelo)
```

```
# Validación de condiciones para la regresión múltiple lineal
```

```
library(ggplot2)
```

```
library(gridExtra)
```

```
# Relación lineal entre los predictores numéricos y la variable dependiente
```

```
plot1 <- ggplot(data = Monash_O3_reg, aes(Temperatura, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
  theme_bw()
```

```
plot2 <- ggplot(data = Monash_O3_reg, aes(Precipitacion, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
```

```

theme_bw()

plot1 <- plot1 + theme(axis.text = element_text(size=12), axis.title = element_text(size =
16))
plot2 <- plot2 + theme(axis.text = element_text(size=12), axis.title = element_text(size =
16))

grid.arrange(plot1, plot2)

# Normalidad
qqnorm(modelo$residuals, main="Grafico QQnorm", ylab= "Cuantiles
muestrales",xlab="Cuantiles teóricos N(0,1)",cex.lab=1.3,cex.axis=1.2)
qqline(modelo$residuals)

# Test de normalidad
ks.test(modelo$residuals, "pnorm", mean = mean(modelo$residuals), sd =
sd(modelo$residuals))

# Variabilidad constante de los residuos:
plot1 <- ggplot(data = data.frame(predict_values = predict(modelo),
residuos = residuals(modelo)),
aes(x = predict_values, y = residuos)) +
geom_point() +
geom_smooth(color = "firebrick", se = FALSE) +
geom_hline(yintercept = 0) +
theme_bw()
plot1 <- plot1 + theme(axis.text = element_text(size=12),axis.title.x = element_text(size
= 16), axis.title.y = element_text(size = 16))+ labs(x='Valores predichos', y='Residuos')
plot1

# Test de homocedasticidad
library(lmtest)
bptest(modelo)

```

```

# Identificación de posibles valores atípicos o influyentes
library(car)
outlierTest(modelo)

summary(influence.measures(modelo))

influencePlot(modelo)

# los que ocupan las posiciones 330,874,889,1322,1884

# Se decide rehacer el modelo excluyendo los outliers y los puntos influyentes

Monash_O3_reg <- Monash_O3_reg[-c(874,889,1615,1257,1615,1884), ]

# Modelo General
modelo <- lm(log(Promedio_O3) ~ Temperatura + Precipitacion + Incendio + Fase.0 +
             Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
Monash_O3_reg )
summary(modelo)

# Selección de los mejores predictores
step(object = modelo, direction = "both", trace = 1)

# Mejor modelo sin puntos influyentes
modelo <- lm(log(Promedio_O3) ~ Temperatura + Precipitacion +
             Fase.1 + Confinamiento + Desconfinamiento + Sabado, data =
Monash_O3_reg)
summary(modelo)

# MONASH - PM10
# Leyendo la nueva data con las variables independientes agregadas

```

```

Monash_PM10_reg=read.csv("Monash_PM10.csv",header = TRUE, sep = ";")

str(Monash_PM10_reg)
Monash_PM10_reg$Fecha <- as.Date(Monash_PM10_reg$Fecha, format = "
%d/%m/%Y")

#Verificacion de datos faltantes de las variables a adidas
miss_var_summary(Monash_PM10_reg)

#Imputacion de Precipitacion y temperatura
imputed_Precip_Temp <- mice(Monash_PM10_reg%>%select('Precipitacion',
'Temperatura'), method = "norm")

# Se asigna a una variable cada imputacion
imputed_Precip_Temp <- mice::complete(imputed_Precip_Temp)

par(mfrow=c(1,2))
plot(density(Monash_PM10_reg$Precipitacion,na.rm = T),col=2,main="Precipitacion")
lines(density(as.numeric(imputed_Precip_Temp$Precipitacion)),col=3)
plot(density(Monash_PM10_reg$Temperatura,na.rm = T),col=2,main="Temperatura")
lines(density(as.numeric(imputed_Precip_Temp$Temperatura)),col=3)

# Se reemplaza las variables imputadas en la base original.
Monash_PM10_reg<-Monash_PM10_reg%>%mutate(Precipitacion =
imputed_Precip_Temp$Precipitacion,
      Temperatura = imputed_Precip_Temp$Temperatura)

# Asignando el tipo de dato que le corresponde
Monash_PM10_reg$Incendio <- as.factor(Monash_PM10_reg$Incendio)
Monash_PM10_reg$Fase.0 <- as.factor(Monash_PM10_reg$Fase.0)
Monash_PM10_reg$Fase.1 <- as.factor(Monash_PM10_reg$Fase.1)
Monash_PM10_reg$Confinamiento <- as.factor(Monash_PM10_reg$Confinamiento)
Monash_PM10_reg$Desconfinamiento <-
as.factor(Monash_PM10_reg$Desconfinamiento)

```

```

Monash_PM10_reg$Sabado <- as.factor(Monash_PM10_reg$Sabado)
Monash_PM10_reg$Domingo <- as.factor(Monash_PM10_reg$Domingo)

# Visualizando las nuevas variables predictoras tipo factor
library(ggpubr) #para gestionar multiples graficos
g1 <- ggplot(data = Monash_PM10_reg, mapping=aes(x = Incendio, y =
Promedio_PM10, color=Incendio)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g2 <- ggplot(data = Monash_PM10_reg, mapping=aes(x = Fase.0, y =
Promedio_PM10, color=Fase.0)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g3 <- ggplot(data = Monash_PM10_reg, mapping=aes(x = Fase.1, y =
Promedio_PM10, color=Fase.1)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g4 <- ggplot(data = Monash_PM10_reg, mapping=aes(x = Confinamiento, y =
Promedio_PM10, color=Confinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g5 <- ggplot(data = Monash_PM10_reg, mapping=aes(x = Desconfinamiento, y =
Promedio_PM10, color=Desconfinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g6 <- ggplot(data = Monash_PM10_reg, mapping=aes(x = Sabado, y =
Promedio_PM10, color=Sabado)) +
  geom_boxplot() +

```

```

geom_jitter(width = 0.1) +
theme_bw() + theme(legend.position = "none")
g7 <- ggplot(data = Monash_PM10_reg, mapping=aes(x = Domingo, y =
Promedio_PM10, color=Domingo)) +
geom_boxplot() +
geom_jitter(width = 0.1) +
theme_bw() + theme(legend.position = "none")

annotate_figure(ggarrange(g1,g2,g3,g4,g5,g6,g7,
ncol = 4, nrow = 2),
top=text_grob("BoxPlot de las variables tipo factor"))

# Regresion
# install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)

chart.Correlation(Monash_PM10_reg[,c('Promedio_PM10','Temperatura',
'Precipitacion')], histogram = TRUE, method = "pearson")

# Modelo General
modelo <- lm(log(Promedio_PM10) ~ Temperatura + Precipitacion + Incendio + Fase.0
+
Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
Monash_PM10_reg )
summary(modelo)

# Validación de condiciones para la regresión múltiple lineal

library(ggplot2)
library(gridExtra)

# Relación lineal entre los predictores numéricos y la variable dependiente
plot1 <- ggplot(data = Monash_PM10_reg, aes(Temperatura, modelo$residuals)) +

```

```

geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
theme_bw()
plot2 <- ggplot(data = Monash_PM10_reg, aes(Precipitacion, modelo$residuals)) +
geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
theme_bw()

plot1 <- plot1 + theme(axis.text = element_text(size=12), axis.title = element_text(size =
16))
plot2 <- plot2 + theme(axis.text = element_text(size=12), axis.title = element_text(size =
16))

grid.arrange(plot1, plot2)

# Normalidad
qqnorm(modelo$residuals, main="Grafico QQnorm", ylab= "Cuantiles
muestrales", xlab="Cuantiles teóricos N(0,1)", cex.lab=1.3, cex.axis=1.2)
qqline(modelo$residuals)

# Test de normalidad
ks.test(modelo$residuals, "pnorm", mean = mean(modelo$residuals), sd =
sd(modelo$residuals))

# Homocedasticidad
# Variabilidad constante de los residuos:
plot1 <- ggplot(data = data.frame(predict_values = predict(modelo),
residuos = residuals(modelo)),
aes(x = predict_values, y = residuos)) +
geom_point() +
geom_smooth(color = "firebrick", se = FALSE) +
geom_hline(yintercept = 0) +
theme_bw()
plot1 <- plot1 + theme(axis.text = element_text(size=12), axis.title.x = element_text(size
= 16), axis.title.y = element_text(size = 16)) + labs(x='Valores predichos', y='Residuos')
plot1

```

```

# Test de homocedasticidad
library(lmtest)
bptest(modelo)

# Identificación de posibles valores atípicos o influyentes
library(car)
outlierTest(modelo)

summary(influence.measures(modelo))

influencePlot(modelo)

# los que ocupan las posiciones 330,1100,1101,1543,1884

# Se decide rehacer el modelo excluyendo los outliers y los puntos influyentes

Monash_PM10_reg <- Monash_PM10_reg[-c(1100,1101,1097,535,330,1218,1884), ]

# Modelo General
modelo <- lm(log(Promedio_PM10) ~ Temperatura + Precipitacion + Incendio + Fase.0
+
      Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
Monash_PM10_reg )
summary(modelo)

# Selección de los mejores predictores
step(object = modelo, direction = "both", trace = 1)

# Mejor modelo sin puntos influyentes
modelo <- lm(log(Promedio_PM10) ~ Temperatura + Precipitacion +
      Incendio + Fase.1 + Desconfinamiento, data = Monash_PM10_reg)
summary(modelo)

```



```

#-----
# Florey
#-----

# Imputacion de los valores faltantes
#-----

library(mice) # permite realizar la imputacion de los datos mediante diferentes técnicas
library(VIM) # permite visualizar los datos faltantes
library(naniar) # Muestra el total de valores faltantes

# Metodo con funcion naniar
miss_var_summary(Florey)

# El metodo norm.boot, realiza la imputacion con regresiones y de esa manera fortalece
por simulacion la relacion en los datos
# Se procedera a imputar agrupando variables, por conveniencia, segun su escala
imputed_data_0 <- mice(Florey%>%select('NO2', 'O3_1hr', 'PM10'), method =
"norm")
imputed_data_1 <- mice(Florey%>%select('AQI_NO2', 'AQI_O3_1hr', 'AQI_PM10',
'AQI_Site'), method = "norm.boot")

# Se asigna a una variable cada imputacion
Data_Impu_0 <- mice::complete(imputed_data_0)
Data_Impu_1 <- mice::complete(imputed_data_1)

#Color rojo es el valor verdadero y verde es la imputacion
par(mfrow=c(1,2))
plot(density(Florey$NO2,na.rm = T),col=2,main="NO2")
lines(density(as.numeric(Data_Impu_0$NO2)),col=3)
plot(density(Florey$O3_1hr,na.rm = T),col=2,main="O3_1hr")

```

```

lines(density(as.numeric(Data_Impu_0$O3_1hr)),col=3)

par(mfrow=c(1,2))
plot(density(Florey$PM10,na.rm = T),col=2,main="PM10")
lines(density(as.numeric(Data_Impu_0$PM10)),col=3)
plot(density(Florey$AQI_NO2,na.rm = T),col=2,main="AQI_NO2")
lines(density(as.numeric(Data_Impu_1$AQI_NO2)),col=3)

par(mfrow=c(1,2))
plot(density(Florey$AQI_O3_1hr,na.rm = T),col=2,main="AQI_O3_1hr")
lines(density(as.numeric(Data_Impu_1$AQI_O3_1hr)),col=3)
plot(density(Florey$AQI_PM10,na.rm = T),col=2,main="AQI_PM10")
lines(density(as.numeric(Data_Impu_1$AQI_PM10)),col=3)

plot(density(Florey$AQI_Site,na.rm = T),col=2,main="AQI_Site")
lines(density(as.numeric(Data_Impu_1$AQI_Site)),col=3)

# Se reemplaza las variables imputadas en la base original.
Florey<-Florey%>%mutate(NO2 = Data_Impu_0$NO2, O3_1hr =
Data_Impu_0$O3_1hr, PM10 = Data_Impu_0$PM10
, AQI_NO2 = Data_Impu_1$AQI_NO2, AQI_O3_1hr =
Data_Impu_1$AQI_O3_1hr
, AQI_PM10 = Data_Impu_1$AQI_PM10, AQI_Site =
Data_Impu_1$AQI_Site)

# Se comprueba que no quedan valores faltantes
miss_var_summary(Florey)

# Extraemos la data para cada particula NO2, O3 y PM10 de Monash
# Agrupamos la data por promedio de particula de dia, mes y año de los periodos 2017-
2022
Florey_NO2 <- Florey %>% mutate(year = format(DateT, "% Y"), month =
format(DateT, "%m"), day = format(DateT, "%d")) %>%

```

```

group_by(year, month, day) %>%
summarize(mean_NO2 = mean(NO2))

Florey_O3 <- Florey %>% mutate(year = format(DateT, "%Y"), month =
format(DateT, "%m"), day = format(DateT, "%d")) %>%
group_by(year, month, day) %>%
summarize(mean_O3 = mean(O3_1hr))

Florey_PM10 <- Florey %>% mutate(year = format(DateT, "%Y"), month =
format(DateT, "%m"), day = format(DateT, "%d")) %>%
group_by(year, month, day) %>%
summarize(mean_PM10 = mean(PM10))

# Convierte de tibble a DataFrame
Florey_NO2 <- Florey_NO2 %>% as.data.frame()
Florey_O3 <- Florey_O3 %>% as.data.frame()
Florey_PM10 <- Florey_PM10 %>% as.data.frame()

#Exportamos a excel para unir con las otras variables dependientes
library("writexl")
write_xlsx(Florey_NO2,"Florey_NO2.xlsx")
write_xlsx(Florey_O3,"Florey_O3.xlsx")
write_xlsx(Florey_PM10,"Florey_PM10.xlsx")

# FLOREY - NO2
# Leyendo la nueva data con las variables independientes agregadas
Florey_NO2_reg=read.csv("Florey_NO2.csv",header = TRUE, sep = ";")

str(Florey_NO2_reg)
Florey_NO2_reg$Fecha <- as.Date(Florey_NO2_reg$Fecha, format = "%d/%m/%Y")

#Verificacion de datos faltantes de las variables añadidas

```

```

miss_var_summary(Florey_NO2_reg)

#Imputacion de Precipitacion y temperatura
# Utilizamos la regresion lineal Lasso que es la que mejor ajusta a los datos
imputed_Precip_Temp <- mice(Florey_NO2_reg%>%select('Precipitacion',
'Temperatura'), method = "lasso.norm")

# Se asigna a una variable cada imputacion
imputed_Precip_Temp <- mice::complete(imputed_Precip_Temp)

par(mfrow=c(1,2))
plot(density(Florey_NO2_reg$Precipitacion,na.rm = T),col=2,main="Precipitacion")
lines(density(as.numeric(imputed_Precip_Temp$Precipitacion)),col=3)
plot(density(Florey_NO2_reg$Temperatura,na.rm = T),col=2,main="Temperatura")
lines(density(as.numeric(imputed_Precip_Temp$Temperatura)),col=3)

# Se reemplaza las variables imputadas en la base original.
Florey_NO2_reg<-Florey_NO2_reg%>%mutate(Precipitacion =
imputed_Precip_Temp$Precipitacion,
          Temperatura = imputed_Precip_Temp$Temperatura)

# Asignando el tipo de dato que le corresponde
Florey_NO2_reg$Incendio <- as.factor(Florey_NO2_reg$Incendio)
Florey_NO2_reg$Fase.0 <- as.factor(Florey_NO2_reg$Fase.0)
Florey_NO2_reg$Fase.1 <- as.factor(Florey_NO2_reg$Fase.1)
Florey_NO2_reg$Confinamiento <- as.factor(Florey_NO2_reg$Confinamiento)
Florey_NO2_reg$Desconfinamiento <- as.factor(Florey_NO2_reg$Desconfinamiento)
Florey_NO2_reg$Sabado <- as.factor(Florey_NO2_reg$Sabado)
Florey_NO2_reg$Domingo <- as.factor(Florey_NO2_reg$Domingo)

# Visualizando las nuevas variables predictoras tipo factor
library(ggpubr) #para gestionar multiples graficos

```

```

g1 <- ggplot(data = Florey_NO2_reg, mapping=aes(x = Incendio, y = Promedio_NO2,
color=Incendio)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g2 <- ggplot(data = Florey_NO2_reg, mapping=aes(x = Fase.0, y = Promedio_NO2,
color=Fase.0)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g3 <- ggplot(data = Florey_NO2_reg, mapping=aes(x = Fase.1, y = Promedio_NO2,
color=Fase.1)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g4 <- ggplot(data = Florey_NO2_reg, mapping=aes(x = Confinamiento, y =
Promedio_NO2, color=Confinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g5 <- ggplot(data = Florey_NO2_reg, mapping=aes(x = Desconfinamiento, y =
Promedio_NO2, color=Desconfinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g6 <- ggplot(data = Florey_NO2_reg, mapping=aes(x = Sabado, y = Promedio_NO2,
color=Sabado)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g7 <- ggplot(data = Florey_NO2_reg, mapping=aes(x = Domingo, y = Promedio_NO2,
color=Domingo)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +

```

```

theme_bw() + theme(legend.position = "none")

annotate_figure(ggarrange(g1,g2,g3,g4,g5,g6,g7,
                        ncol = 4, nrow = 2),
               top=text_grob("BoxPlot de las variables tipo factor"))

# Regresion

# install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)

chart.Correlation(Florey_NO2_reg[,c('Promedio_NO2','Temperatura', 'Precipitacion')],
                 histogram = TRUE, method = "pearson")

# Modelo General
modelo <- lm(Promedio_NO2 ~ Temperatura + Precipitacion + Incendio + Fase.0 +
            Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
            Florey_NO2_reg )
summary(modelo)

# Validación de condiciones para la regresión múltiple lineal

library(ggplot2)
library(gridExtra)

# Relación lineal entre los predictores numéricos y la variable dependiente
plot1 <- ggplot(data = Florey_NO2_reg, aes(Temperatura, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
  theme_bw()
plot2 <- ggplot(data = Florey_NO2_reg, aes(Precipitacion, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
  theme_bw()

```

```
plot1 <- plot1+ theme(axis.text = element_text(size=12), axis.title = element_text(size = 16))
```

```
plot2 <- plot2+ theme(axis.text = element_text(size=12), axis.title = element_text(size = 16))
```

```
grid.arrange(plot1, plot2)
```

```
# Normalidad
```

```
qqnorm(modelo$residuals, main="Grafico QQnorm", ylab= "Cuantiles muestrales",xlab="Cuantiles teóricos N(0,1)",cex.lab=1.3,cex.axis=1.2)  
qqline(modelo$residuals)
```

```
# Homocedasticidad
```

```
ks.test(modelo$residuals, "pnorm", mean = mean(modelo$residuals), sd = sd(modelo$residuals))
```

```
# Variabilidad constante de los residuos:
```

```
plot1 <-ggplot(data = data.frame(predict_values = predict(modelo),  
                                residuos = residuals(modelo)),  
              aes(x = predict_values, y = residuos)) +  
  geom_point() +  
  geom_smooth(color = "firebrick", se = FALSE) +  
  geom_hline(yintercept = 0) +  
  theme_bw()
```

```
plot1 <- plot1 + theme(axis.text = element_text(size=12),axis.title.x = element_text(size = 16), axis.title.y = element_text(size = 16))+ labs(x='Valores predichos', y='Residuos')
```

```
plot1
```

```
library(lmtest)
```

```
bptest(modelo)
```

```

# Identificación de posibles valores atípicos o influyentes
library(car)
outlierTest(modelo)

summary(influence.measures(modelo))

influencePlot(modelo)

# los que ocupan las posiciones 422,750,1038,1136

# Se decide rehacer el modelo excluyendo los outliers y los puntos influyentes

Florey_NO2_reg <- Florey_NO2_reg[-c(1100,994,987,1096,993,422,1136), ]

# Modelo General
modelo <- lm(Promedio_NO2 ~ Temperatura + Precipitacion + Incendio + Fase.0 +
             Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
Florey_NO2_reg )
summary(modelo)

# Selección de los mejores predictores
step(object = modelo, direction = "both", trace = 1)

# Mejor modelo sin puntos influyentes
modelo <- lm(log(Promedio_NO2+0.01) ~ Temperatura + Precipitacion + Incendio +
             Fase.0 + Fase.1 + Confinamiento + Desconfinamiento + Sabado +
             Domingo, data = Florey_NO2_reg)
summary(modelo)

# FLOREY - O3
# Leyendo la nueva data con las variables independientes agregadas
Florey_O3_reg=read.csv("Florey_O3.csv",header = TRUE, sep = ";")

```



```

str(Florey_O3_reg)
Florey_O3_reg$Fecha <- as.Date(Florey_O3_reg$Fecha, format = " %d/%m/%Y")

#Verificacion de datos faltantes de las variables añadidas
miss_var_summary(Florey_O3_reg)

#Imputacion de Precipitacion y temperatura
imputed_Precip_Temp <- mice(Florey_O3_reg%>%select('Precipitacion',
'Temperatura'), method = "norm")

# Se asigna a una variable cada imputacion
imputed_Precip_Temp <- mice::complete(imputed_Precip_Temp)

par(mfrow=c(1,2))
plot(density(Florey_O3_reg$Precipitacion,na.rm = T),col=2,main="Precipitacion")
lines(density(as.numeric(imputed_Precip_Temp$Precipitacion)),col=3)
plot(density(Florey_O3_reg$Temperatura,na.rm = T),col=2,main="Temperatura")
lines(density(as.numeric(imputed_Precip_Temp$Temperatura)),col=3)

# Se reemplaza las variables imputadas en la base original.
Florey_O3_reg<-Florey_O3_reg%>%mutate(Precipitacion =
imputed_Precip_Temp$Precipitacion,
          Temperatura = imputed_Precip_Temp$Temperatura)

# Asignando el tipo de dato que le corresponde
Florey_O3_reg$Incendio <- as.factor(Florey_O3_reg$Incendio)
Florey_O3_reg$Fase.0 <- as.factor(Florey_O3_reg$Fase.0)
Florey_O3_reg$Fase.1 <- as.factor(Florey_O3_reg$Fase.1)
Florey_O3_reg$Confinamiento <- as.factor(Florey_O3_reg$Confinamiento)
Florey_O3_reg$Desconfinamiento <- as.factor(Florey_O3_reg$Desconfinamiento)
Florey_O3_reg$Sabado <- as.factor(Florey_O3_reg$Sabado)
Florey_O3_reg$Domingo <- as.factor(Florey_O3_reg$Domingo)

# Visualizando las nuevas variables predictoras tipo factor

```

```

library(ggpubr) #para gestionar multiples graficos
g1 <- ggplot(data = Florey_O3_reg, mapping=aes(x = Incendio, y = Promedio_O3,
color=Incendio)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g2 <- ggplot(data = Florey_O3_reg, mapping=aes(x = Fase.0, y = Promedio_O3,
color=Fase.0)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g3 <- ggplot(data = Florey_O3_reg, mapping=aes(x = Fase.1, y = Promedio_O3,
color=Fase.1)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g4 <- ggplot(data = Florey_O3_reg, mapping=aes(x = Confinamiento, y =
Promedio_O3, color=Confinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g5 <- ggplot(data = Florey_O3_reg, mapping=aes(x = Desconfinamiento, y =
Promedio_O3, color=Desconfinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g6 <- ggplot(data = Florey_O3_reg, mapping=aes(x = Sabado, y = Promedio_O3,
color=Sabado)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g7 <- ggplot(data = Florey_O3_reg, mapping=aes(x = Domingo, y = Promedio_O3,
color=Domingo)) +
  geom_boxplot() +

```

```

geom_jitter(width = 0.1) +
theme_bw() + theme(legend.position = "none")

annotate_figure(ggarrange(g1,g2,g3,g4,g5,g6,g7,
                        ncol = 4, nrow = 2),
                top=text_grob("BoxPlot de las variables tipo factor"))

# Regresion
# install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)

chart.Correlation(Florey_O3_reg[,c('Promedio_O3','Temperatura', 'Precipitacion')],
                 histogram = TRUE, method = "pearson")

# Modelo General
modelo <- lm(Promedio_O3 ~ Temperatura + Precipitacion + Incendio + Fase.0 +
            Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
            Florey_O3_reg )
summary(modelo)

# Validación de condiciones para la regresión múltiple lineal

library(ggplot2)
library(gridExtra)

# Relación lineal entre los predictores numéricos y la variable dependiente
plot1 <- ggplot(data = Florey_O3_reg, aes(Temperatura, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
  theme_bw()
plot2 <- ggplot(data = Florey_O3_reg, aes(Precipitacion, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
  theme_bw()

```

```
plot1 <- plot1+ theme(axis.text = element_text(size=12), axis.title = element_text(size = 16))
```

```
plot2 <- plot2+ theme(axis.text = element_text(size=12), axis.title = element_text(size = 16))
```

```
grid.arrange(plot1, plot2)
```

```
# Normalidad
```

```
qqnorm(modelo$residuals, main="Grafico QQnorm", ylab= "Cuantiles muestrales",xlab="Cuantiles teóricos N(0,1)",cex.lab=1.3,cex.axis=1.2)
```

```
qqline(modelo$residuals)
```

```
ks.test(modelo$residuals, "pnorm", mean = mean(modelo$residuals), sd = sd(modelo$residuals))
```

```
# Homocedasticidad
```

```
# Variabilidad constante de los residuos:
```

```
plot1 <- ggplot(data = data.frame(predict_values = predict(modelo),  
                                residuos = residuals(modelo)),
```

```
  aes(x = predict_values, y = residuos)) +
```

```
  geom_point() +
```

```
  geom_smooth(color = "firebrick", se = FALSE) +
```

```
  geom_hline(yintercept = 0) +
```

```
  theme_bw()
```

```
plot1 <- plot1 + theme(axis.text = element_text(size=12),axis.title.x = element_text(size = 16), axis.title.y = element_text(size = 16))+ labs(x='Valores predichos', y='Residuos')
```

```
plot1
```

```
library(lmtest)
```

```
bptest(modelo)
```

```
# Identificación de posibles valores atípicos o influyentes
```

```
library(car)
```

```

outlierTest(modelo)
# Los que ocupan las posiciones 889,1615

# Tal como se apreció en el estudio de normalidad de los residuos,
# las observacions 889,1615 tiene residuos estandarizado >3 (más de 3 veces la
desviación estándar de los residuos) por lo que se considera un dato atípico.
# El siguiente paso es determinar si es influyente

summary(influence.measures(modelo))

influencePlot(modelo)

# los que ocupan las posiciones 422,889,1136,1322,1615

# Se decide rehacer el modelo excluyendo los outliers y los puntos influyentes

Florey_O3_reg <- Florey_O3_reg[-c(1096,1084,422,1136,1218), ]

# Modelo General
modelo <- lm(Promedio_O3 ~ Temperatura + Precipitacion + Incendio + Fase.0 +
             Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
Florey_O3_reg )
summary(modelo)

# Selección de los mejores predictores
step(object = modelo, direction = "both", trace = 1)

# Mejor modelo sin puntos influyentes
modelo <- lm(log(Promedio_O3) ~ Temperatura + Precipitacion + Incendio +
             Fase.0 + Fase.1 + Confinamiento + Sabado, data = Florey_O3_reg)
summary(modelo)

# FLOREY - PM10

```

```

# Leyendo los nuevos datos con las variables independientes agregadas
Florey_PM10_reg=read.csv("Florey_PM10.csv",header = TRUE, sep = ";")

str(Florey_PM10_reg)
Florey_PM10_reg$Fecha <- as.Date(Florey_PM10_reg$Fecha, format = "
%d/%m/%Y")

#Verificacion de datos faltantes de las variables añadidas
miss_var_summary(Florey_PM10_reg)

#Imputacion de Precipitacion y temperatura
imputed_Precip_Temp <- mice(Florey_PM10_reg%>%select('Precipitacion',
'Temperatura'), method = "rf")

# Se asigna a una variable cada imputacion
imputed_Precip_Temp <- mice::complete(imputed_Precip_Temp)

par(mfrow=c(1,2))
plot(density(Florey_PM10_reg$Precipitacion,na.rm = T),col=2,main="Precipitacion")
lines(density(as.numeric(imputed_Precip_Temp$Precipitacion)),col=3)
plot(density(Florey_PM10_reg$Temperatura,na.rm = T),col=2,main="Temperatura")
lines(density(as.numeric(imputed_Precip_Temp$Temperatura)),col=3)

# Se reemplaza las variables imputadas en la base original.
Florey_PM10_reg<-Florey_PM10_reg%>%mutate(Precipitacion =
imputed_Precip_Temp$Precipitacion,
          Temperatura = imputed_Precip_Temp$Temperatura)

# Asignando el tipo de dato que le corresponde
Florey_PM10_reg$Incendio <- as.factor(Florey_PM10_reg$Incendio)
Florey_PM10_reg$Fase.0 <- as.factor(Florey_PM10_reg$Fase.0)
Florey_PM10_reg$Fase.1 <- as.factor(Florey_PM10_reg$Fase.1)
Florey_PM10_reg$Confinamiento <- as.factor(Florey_PM10_reg$Confinamiento)

```

```

Florey_PM10_reg$Desconfinamiento <-
as.factor(Florey_PM10_reg$Desconfinamiento)
Florey_PM10_reg$Sabado <- as.factor(Florey_PM10_reg$Sabado)
Florey_PM10_reg$Domingo <- as.factor(Florey_PM10_reg$Domingo)

# Visualizando las nuevas variables predictoras tipo factor
library(ggpubr) #para gestionar multiples graficos
g1 <- ggplot(data = Florey_PM10_reg, mapping=aes(x = Incendio, y =
Promedio_PM10, color=Incendio)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g2 <- ggplot(data = Florey_PM10_reg, mapping=aes(x = Fase.0, y = Promedio_PM10,
color=Fase.0)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g3 <- ggplot(data = Florey_PM10_reg, mapping=aes(x = Fase.1, y = Promedio_PM10,
color=Fase.1)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g4 <- ggplot(data = Florey_PM10_reg, mapping=aes(x = Confinamiento, y =
Promedio_PM10, color=Confinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g5 <- ggplot(data = Florey_PM10_reg, mapping=aes(x = Desconfinamiento, y =
Promedio_PM10, color=Desconfinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g6 <- ggplot(data = Florey_PM10_reg, mapping=aes(x = Sabado, y = Promedio_PM10,
color=Sabado)) +

```

```

geom_boxplot() +
geom_jitter(width = 0.1) +
theme_bw() + theme(legend.position = "none")
g7 <- ggplot(data = Florey_PM10_reg, mapping=aes(x = Domingo, y =
Promedio_PM10, color=Domingo)) +
geom_boxplot() +
geom_jitter(width = 0.1) +
theme_bw() + theme(legend.position = "none")

annotate_figure(ggarrange(g1,g2,g3,g4,g5,g6,g7,
ncol = 4, nrow = 2),
top=text_grob("BoxPlot de las variables tipo factor"))

# Regresion
# install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)

chart.Correlation(Florey_PM10_reg[,c('Promedio_PM10','Temperatura',
'Precipitacion')], histogram = TRUE, method = "pearson")

# Modelo General
modelo <- lm(Promedio_PM10 ~ Temperatura + Precipitacion + Incendio + Fase.0 +
Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
Florey_PM10_reg )
summary(modelo)

# Validación de condiciones para la regresión múltiple lineal
library(ggplot2)
library(gridExtra)

# Relación lineal entre los predictores numéricos y la variable dependiente
plot1 <- ggplot(data = Florey_PM10_reg, aes(Temperatura, modelo$residuals)) +
geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +

```



```

theme_bw()
plot2 <- ggplot(data = Florey_PM10_reg, aes(Precipitacion, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
  theme_bw()
plot1 <- plot1 + theme(axis.text = element_text(size=12), axis.title = element_text(size =
16))
plot2 <- plot2 + theme(axis.text = element_text(size=12), axis.title = element_text(size =
16))

grid.arrange(plot1, plot2)

# Normalidad
qqnorm(modelo$residuals, main="Grafico QQnorm", ylab= "Cuantiles
muestrales", xlab="Cuantiles teóricos N(0,1)", cex.lab=1.3, cex.axis=1.2)
qqline(modelo$residuals)

ks.test(modelo$residuals, "pnorm", mean = mean(modelo$residuals), sd =
sd(modelo$residuals))

# Homocedasticidad
# Variabilidad constante de los residuos:
plot1 <- ggplot(data = data.frame(predict_values = predict(modelo),
  residuos = residuals(modelo)),
  aes(x = predict_values, y = residuos)) +
  geom_point() +
  geom_smooth(color = "firebrick", se = FALSE) +
  geom_hline(yintercept = 0) +
  theme_bw()
plot1 <- plot1 + theme(axis.text = element_text(size=12), axis.title.x = element_text(size
= 16), axis.title.y = element_text(size = 16)) + labs(x='Valores predichos', y='Residuos')
plot1

library(lmtest)
bptest(modelo)

```

```

# Identificación de posibles valores atípicos o influyentes
library(car)
outlierTest(modelo)

summary(influence.measures(modelo))

influencePlot(modelo)

# los que ocupan las posiciones 442,1096,1097,1136

# Se decide rehacer el modelo excluyendo los outliers y los puntos influyentes

Florey_PM10_reg <- Florey_PM10_reg[-
c(1096,1097,1100,1101,1085,1098,1086,1084,774,1082,442,1136), ]

# Modelo General
modelo <- lm(Promedio_PM10 ~ Temperatura + Precipitacion + Incendio + Fase.0 +
             Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
Florey_PM10_reg )
summary(modelo)

# Selección de los mejores predictores
step(object = modelo, direction = "both", trace = 1)

# Mejor modelo sin puntos influyentes
modelo <- lm(log(Promedio_PM10+0.01) ~ Temperatura + Precipitacion + Incendio +
             Fase.0 + Fase.1 + Sabado, data = Florey_PM10_reg)
summary(modelo)

#-----

```

```

# Civic
#-----

# Imputacion de los valores faltantes
#-----

library(mice) # permite realizar la imputacion de los datos mediante diferentes técnicas
library(VIM) # permite visualizar los datos faltantes
library(naniar) # Muestra el total de valores faltantes

# Metodo con funcion naniar
miss_var_summary(Civic)

#Descartamos imputar NO2 y AQINO2
Civic <- Civic%>% select (-one_of ('NO2', 'AQI_NO2'))

# El metodo norm.boot, realiza la imputacion con regresiones y de esa manera fortalece
por simulacion la relacion en los datos
# Se procedera a imputar agrupando variables, por conveniencia, segun su escala
imputed_data_0 <- mice(Civic%>%select('O3_1hr', 'PM10'), method = "norm")
imputed_data_1 <- mice(Civic%>%select('AQI_O3_1hr', 'AQI_PM10', 'AQI_Site'),
method = "norm.boot")

# Se asigna a una variable cada imputacion
Data_Impu_0 <- mice::complete(imputed_data_0)
Data_Impu_1 <- mice::complete(imputed_data_1)

#Color rojo es el valor verdadero y verde es la imputacion
par(mfrow=c(1,2))
plot(density(Civic$O3_1hr,na.rm = T),col=2,main="O3_1hr")
lines(density(as.numeric(Data_Impu_0$O3_1hr)),col=3)
plot(density(Civic$PM10,na.rm = T),col=2,main="PM10")
lines(density(as.numeric(Data_Impu_0$PM10)),col=3)

```

```

par(mfrow=c(1,2))
plot(density(Civic$AQI_O3_1hr,na.rm = T),col=2,main="AQI_O3_1hr")
lines(density(as.numeric(Data_Impu_1$AQI_O3_1hr)),col=3)
plot(density(Civic$AQI_PM10,na.rm = T),col=2,main="AQI_PM10")
lines(density(as.numeric(Data_Impu_1$AQI_PM10)),col=3)

plot(density(Civic$AQI_Site,na.rm = T),col=2,main="AQI_Site")
lines(density(as.numeric(Data_Impu_1$AQI_Site)),col=3)

# Se reemplaza las variables imputadas en la base original.
Civic<-Civic%>% mutate(O3_1hr = Data_Impu_0$O3_1hr, PM10 =
Data_Impu_0$PM10
, AQI_O3_1hr = Data_Impu_1$AQI_O3_1hr
, AQI_PM10 = Data_Impu_1$AQI_PM10, AQI_Site =
Data_Impu_1$AQI_Site)

# Se comprueba que no quedan valores faltantes
miss_var_summary(Civic)

# Extraemos la data para cada particula NO2, O3 y PM10 de Monash
# Agrupamos la data por promedio de particula de dia, mes y año de los periodos 2017-
2022

Civic_O3 <- Civic %>% mutate(year = format(DateT, "%Y"), month = format(DateT,
"%m"), day = format(DateT, "%d")) %>%
group_by(year, month, day) %>%
summarize(mean_O3 = mean(O3_1hr))

Civic_PM10 <- Civic %>% mutate(year = format(DateT, "%Y"), month =
format(DateT, "%m"), day = format(DateT, "%d")) %>%
group_by(year, month, day) %>%
summarize(mean_PM10 = mean(PM10))

```

```

# Convierte de tibble a DataFrame
Civic_O3 <- Civic_O3 %>% as.data.frame()
Civic_PM10 <- Civic_PM10 %>% as.data.frame()

summary(Civic)

#Exportamos a excel para unir con las otras variables dependientes
library("writexl")
write_xlsx(Civic_O3,"Civic_O3.xlsx")
write_xlsx(Civic_PM10,"Civic_PM10.xlsx")

# CIVIC - O3
# Leyendo la nueva data con las variables independientes agregadas
Civic_O3_reg=read.csv("Civic_O3.csv",header = TRUE, sep = ";")

str(Civic_O3_reg)
Civic_O3_reg$Fecha <- as.Date(Civic_O3_reg$Fecha, format = " %d/%m/% Y")

#Verificacion de datos faltantes de las variables añadidas
miss_var_summary(Civic_O3_reg)

#Imputacion de Precipitacion y temperatura
imputed_Precip_Temp <- mice(Civic_O3_reg%>%select('Precipitacion',
'Temperatura'), method = "rf")

# Se asigna a una variable cada imputacion
imputed_Precip_Temp <- mice::complete(imputed_Precip_Temp)

par(mfrow=c(1,2))
plot(density(Civic_O3_reg$Precipitacion,na.rm = T),col=2,main="Precipitacion")
lines(density(as.numeric(imputed_Precip_Temp$Precipitacion)),col=3)
plot(density(Civic_O3_reg$Temperatura,na.rm = T),col=2,main="Temperatura")
lines(density(as.numeric(imputed_Precip_Temp$Temperatura)),col=3)

```

```

# Se reemplaza las variables imputadas en la base original.
Civic_O3_reg<-Civic_O3_reg%>%mutate(Precipitacion =
imputed_Precip_Temp$Precipitacion,
                                Temperatura = imputed_Precip_Temp$Temperatura)

# Asignando el tipo de dato que le corresponde
Civic_O3_reg$Incendio <- as.factor(Civic_O3_reg$Incendio)
Civic_O3_reg$Fase.0 <- as.factor(Civic_O3_reg$Fase.0)
Civic_O3_reg$Fase.1 <- as.factor(Civic_O3_reg$Fase.1)
Civic_O3_reg$Confinamiento <- as.factor(Civic_O3_reg$Confinamiento)
Civic_O3_reg$Desconfinamiento <- as.factor(Civic_O3_reg$Desconfinamiento)
Civic_O3_reg$Sabado <- as.factor(Civic_O3_reg$Sabado)
Civic_O3_reg$Domingo <- as.factor(Civic_O3_reg$Domingo)

# Visualizando las nuevas variables predictoras tipo factor
library(ggpubr) #para gestionar multiples graficos
g1 <- ggplot(data = Civic_O3_reg, mapping=aes(x = Incendio, y = Promedio_O3,
color=Incendio)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g2 <- ggplot(data = Civic_O3_reg, mapping=aes(x = Fase.0, y = Promedio_O3,
color=Fase.0)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g3 <- ggplot(data = Civic_O3_reg, mapping=aes(x = Fase.1, y = Promedio_O3,
color=Fase.1)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")

```

```

g4 <- ggplot(data = Civic_O3_reg, mapping=aes(x = Confinamiento, y = Promedio_O3,
color=Confinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g5 <- ggplot(data = Civic_O3_reg, mapping=aes(x = Desconfinamiento, y =
Promedio_O3, color=Desconfinamiento)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g6 <- ggplot(data = Civic_O3_reg, mapping=aes(x = Sabado, y = Promedio_O3,
color=Sabado)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g7 <- ggplot(data = Civic_O3_reg, mapping=aes(x = Domingo, y = Promedio_O3,
color=Domingo)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")

annotate_figure(ggarrange(g1,g2,g3,g4,g5,g6,g7,
  ncol = 4, nrow = 2),
  top=text_grob("BoxPlot de las variables tipo factor"))

```

```
# Regresion
```

```
# install.packages("PerformanceAnalytics")
```

```
library(PerformanceAnalytics)
```

```
chart.Correlation(Civic_O3_reg[,c('Promedio_O3','Temperatura', 'Precipitacion')],
  histogram = TRUE, method = "pearson")
```

```

# Modelo General
modelo <- lm(Promedio_O3 ~ Temperatura + Precipitacion + Incendio + Fase.0 +
             Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
             Civic_O3_reg )
summary(modelo)

# Validación de condiciones para la regresión múltiple lineal
library(ggplot2)
library(gridExtra)

# Relación lineal entre los predictores numéricos y la variable dependiente
plot1 <- ggplot(data = Civic_O3_reg, aes(Temperatura, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
  theme_bw()
plot2 <- ggplot(data = Civic_O3_reg, aes(Precipitacion, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
  theme_bw()

plot1 <- plot1+ theme(axis.text = element_text(size=12), axis.title = element_text(size =
16))
plot2 <- plot2+ theme(axis.text = element_text(size=12), axis.title = element_text(size =
16))

grid.arrange(plot1, plot2)

# Normalidad
qqnorm(modelo$residuals, main="Grafico QQnorm", ylab= "Cuantiles
muestrales",xlab="Cuantiles teóricos N(0,1)",cex.lab=1.3,cex.axis=1.2)
qqline(modelo$residuals)

ks.test(modelo$residuals, "pnorm", mean = mean(modelo$residuals), sd =
sd(modelo$residuals))

# Homocedasticidad

```



```

# Variabilidad constante de los residuos:
plot1 <- ggplot(data = data.frame(predict_values = predict(modelo),
                                residuos = residuals(modelo)),
               aes(x = predict_values, y = residuos)) +
  geom_point() +
  geom_smooth(color = "firebrick", se = FALSE) +
  geom_hline(yintercept = 0) +
  theme_bw()
plot1 <- plot1 + theme(axis.text = element_text(size=12),axis.title.x = element_text(size
= 16), axis.title.y = element_text(size = 16))+ labs(x='Valores predichos', y='Residuos')
plot1

library(lmtest)
bptest(modelo)

# Identificación de posibles valores atípicos o influyentes
library(car)
outlierTest(modelo)

summary(influence.measures(modelo))

influencePlot(modelo)

# los que ocupan las posiciones 338,422,1062,1067,1289

# Se decide rehacer el modelo excluyendo los outliers y los puntos influyentes

Civic_O3_reg <- Civic_O3_reg[-c(1212,422,1062,1067,1289), ]

# Modelo General
modelo <- lm(Promedio_O3 ~ Temperatura + Precipitacion + Incendio + Fase.0 +
             Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
             Civic_O3_reg )
summary(modelo)

```

```

# Selección de los mejores predictores
step(object = modelo, direction = "both", trace = 1)

# Mejor modelo sin puntos influyentes
modelo <- lm(log(Promedio_O3+0.1) ~ Temperatura + Precipitacion + Incendio +
             Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
             Civic_O3_reg)
summary(modelo)

# CIVIC - PM10
# Leyendo los nuevos datos con las variables independientes agregadas
Civic_PM10_reg=read.csv("Civic_PM10.csv",header = TRUE, sep = ";")

str(Civic_PM10_reg)
Civic_PM10_reg$Fecha <- as.Date(Civic_PM10_reg$Fecha, format = " %d/%m/%Y")

#Verificacion de datos faltantes de las variables añadidas
miss_var_summary(Civic_PM10_reg)

#Imputacion de Precipitacion y temperatura
imputed_Precip_Temp <- mice(Civic_PM10_reg%>%select('Precipitacion',
'Temperatura'), method = "rf")

# Se asigna a una variable cada imputacion
imputed_Precip_Temp <- mice::complete(imputed_Precip_Temp)

par(mfrow=c(1,2))
plot(density(Civic_PM10_reg$Precipitacion,na.rm = T),col=2,main="Precipitacion")
lines(density(as.numeric(imputed_Precip_Temp$Precipitacion)),col=3)
plot(density(Civic_PM10_reg$Temperatura,na.rm = T),col=2,main="Temperatura")
lines(density(as.numeric(imputed_Precip_Temp$Temperatura)),col=3)

```

```

# Se reemplaza las variables imputadas en la base original.
Civic_PM10_reg<-Civic_PM10_reg%>%mutate(Precipitacion =
imputed_Precip_Temp$Precipitacion,
          Temperatura = imputed_Precip_Temp$Temperatura)

# Asignando el tipo de dato que le corresponde
Civic_PM10_reg$Incendio <- as.factor(Civic_PM10_reg$Incendio)
Civic_PM10_reg$Fase.0 <- as.factor(Civic_PM10_reg$Fase.0)
Civic_PM10_reg$Fase.1 <- as.factor(Civic_PM10_reg$Fase.1)
Civic_PM10_reg$Confinamiento <- as.factor(Civic_PM10_reg$Confinamiento)
Civic_PM10_reg$Desconfinamiento <- as.factor(Civic_PM10_reg$Desconfinamiento)
Civic_PM10_reg$Sabado <- as.factor(Civic_PM10_reg$Sabado)
Civic_PM10_reg$Domingo <- as.factor(Civic_PM10_reg$Domingo)

# Visualizando las nuevas variables predictoras tipo factor
library(ggpubr) #para gestionar multiples graficos
g1 <- ggplot(data = Civic_PM10_reg, mapping=aes(x = Incendio, y = Promedio_PM10,
color=Incendio)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g2 <- ggplot(data = Civic_PM10_reg, mapping=aes(x = Fase.0, y = Promedio_PM10,
color=Fase.0)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g3 <- ggplot(data = Civic_PM10_reg, mapping=aes(x = Fase.1, y = Promedio_PM10,
color=Fase.1)) +
  geom_boxplot() +
  geom_jitter(width = 0.1) +
  theme_bw() + theme(legend.position = "none")
g4 <- ggplot(data = Civic_PM10_reg, mapping=aes(x = Confinamiento, y =
Promedio_PM10, color=Confinamiento)) +

```

```

geom_boxplot() +
geom_jitter(width = 0.1) +
theme_bw() + theme(legend.position = "none")
g5 <- ggplot(data = Civic_PM10_reg, mapping=aes(x = Desconfinamiento, y =
Promedio_PM10, color=Desconfinamiento)) +
geom_boxplot() +
geom_jitter(width = 0.1) +
theme_bw() + theme(legend.position = "none")
g6 <- ggplot(data = Civic_PM10_reg, mapping=aes(x = Sabado, y = Promedio_PM10,
color=Sabado)) +
geom_boxplot() +
geom_jitter(width = 0.1) +
theme_bw() + theme(legend.position = "none")
g7 <- ggplot(data = Civic_PM10_reg, mapping=aes(x = Domingo, y =
Promedio_PM10, color=Domingo)) +
geom_boxplot() +
geom_jitter(width = 0.1) +
theme_bw() + theme(legend.position = "none")

annotate_figure(ggarrange(g1,g2,g3,g4,g5,g6,g7,
ncol = 4, nrow = 2),
top=text_grob("BoxPlot de las variables tipo factor"))

```

```
# Regresion
```

```
# install.packages("PerformanceAnalytics")
```

```
library(PerformanceAnalytics)
```

```
chart.Correlation(Civic_PM10_reg[,c('Promedio_PM10','Temperatura', 'Precipitacion')],
histogram = TRUE, method = "pearson")
```

```
# Modelo General
```

```
modelo <- lm(Promedio_PM10 ~ Temperatura + Precipitacion + Incendio + Fase.0 +
```

```

Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
Civic_PM10_reg )
summary(modelo)

# Validación de condiciones para la regresión múltiple lineal

library(ggplot2)
library(gridExtra)

# Relación lineal entre los predictores numéricos y la variable dependiente
plot1 <- ggplot(data = Civic_PM10_reg, aes(Temperatura, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
  theme_bw()
plot2 <- ggplot(data = Civic_PM10_reg, aes(Precipitacion, modelo$residuals)) +
  geom_point() + geom_smooth(color = "firebrick") + geom_hline(yintercept = 0) +
  theme_bw()
plot1 <- plot1 + theme(axis.text = element_text(size=12), axis.title = element_text(size =
16))
plot2 <- plot2 + theme(axis.text = element_text(size=12), axis.title = element_text(size =
16))

grid.arrange(plot1, plot2)

# Normalidad
qqnorm(modelo$residuals, main="Grafico QQnorm", ylab= "Cuantiles
muestrales", xlab="Cuantiles teóricos N(0,1)", cex.lab=1.3, cex.axis=1.2)
qqline(modelo$residuals)

ks.test(modelo$residuals, "pnorm", mean = mean(modelo$residuals), sd =
sd(modelo$residuals))

# Homocedasticidad
# Variabilidad constante de los residuos:

```

```

plot1 <- ggplot(data = data.frame(predict_values = predict(modelo),
                                residuos = residuals(modelo)),
               aes(x = predict_values, y = residuos)) +
  geom_point() +
  geom_smooth(color = "firebrick", se = FALSE) +
  geom_hline(yintercept = 0) +
  theme_bw()
plot1 <- plot1 + theme(axis.text = element_text(size=12),axis.title.x = element_text(size
= 16), axis.title.y = element_text(size = 16))+ labs(x='Valores predichos', y='Residuos')
plot1

```

```

library(lmtest)
bptest(modelo)

```

Identificación de posibles valores atípicos o influyentes

```

library(car)
outlierTest(modelo)

```

```

summary(influence.measures(modelo))

```

```

influencePlot(modelo)

```

los que ocupan las posiciones 442,1067,1068,1071

Se decide rehacer el modelo excluyendo los outliers y los puntos influyentes

```

Civic_PM10_reg <- Civic_PM10_reg[-
c(1067,1068,1071,1072,1056,1069,1055,1057,686,745,442,1067,1289), ]

```

Modelo General

```

modelo <- lm(Promedio_PM10 ~ Temperatura + Precipitacion + Incendio + Fase.0 +
             Fase.1 + Confinamiento + Desconfinamiento + Sabado + Domingo, data =
Civic_PM10_reg )
summary(modelo)

```

```
# Selección de los mejores predictores
step(object = modelo, direction = "both", trace = 1)

# Mejor modelo sin puntos influyentes
modelo <- lm(log(Promedio_PM10+0.1) ~ Temperatura + Precipitacion + Incendio,
             data = Civic_PM10_reg)
summary(modelo)
```