



Tecnológico de Monterrey

Materia:

Análisis de Métodos Multivariados para Ciencia de Datos

Reto - Reporte Final Equipo 2

Profesores:

Blanca Rosa Ruiz Hernández

Monica Guadalupe Elizondo Amaya

Alumnos:

Gian Marco Innocenti A00834310

Andrés Villarreal González A00833915

Andrea Hernandez A00835225

22 de octubre de 2023, Monterrey, N.L.

Resumen de la revisión de bibliografía

El análisis de la calidad del aire es esencial debido a su impacto en la salud y el medio ambiente. La contaminación del aire puede causar graves problemas de salud, como enfermedades respiratorias y cardiovasculares, y contribuir al cambio climático y dañar los ecosistemas. En este análisis, se miden diversos contaminantes comunes, como el ozono, el monóxido de carbono, el dióxido de azufre, el dióxido de nitrógeno y las partículas en suspensión, considerando factores como las condiciones meteorológicas y la ubicación de las fuentes de emisión.

La Organización Mundial de la Salud (OMS) establece directrices globales para la calidad del aire, centrándose en contaminantes clave como las partículas finas, el ozono, el dióxido de nitrógeno, el dióxido de azufre y el monóxido de carbono, debido a su daño potencial para la salud. La calidad del aire se mide mediante el Índice de Calidad del Aire (AQI), que utiliza una escala de 0 a 500 y monitores especializados. Para mejorar la calidad del aire, se pueden reducir las emisiones de contaminantes, mejorar la ventilación y utilizar sistemas de limpieza y filtración de aire. La OMS establece normas oficiales para los niveles de calidad del aire en todo el mundo, pero medir la calidad del aire puede ser un desafío en países en desarrollo debido a la falta de acceso a tecnología avanzada y sistemas de alerta.

En el año 2020 la pandemia del covid 19 cambió el flujo y la movilidad en todas las áreas metropolitanas. Durante ese reporte se buscará verificar si la reducción de la movilidad y la reducción de actividades tanto como civiles e industriales durante este año tuvo un efecto en la concentración de los contaminantes en comparación a 2021 donde la mayoría de actividades económicas y civiles volvieron a la normalidad. Esto se hará con la información recabada por tres estaciones de SIMA. Se escogió una estación en el Norte de la ciudad, otra en el centro de la ciudad, y otra en el sur de la ciudad. Con estas tres estaciones se busco tener información acerca de tres áreas diferentes de la ciudad de Monterrey y ver si se siguieron las mismas tendencias en 2020 y 2021.

Durante este estudio se midieron las concentraciones de cuatro contaminantes para su comparación. Se utilizaron los contaminantes PM10, Co, NO₂, y O₃. Para cada contaminante se buscó un respaldo por medio de datos externos para lograr explicar los resultados de nuestras pruebas estadísticas. Con ayuda de información de terceros logramos respaldar nuestros resultados enfocándonos en información de los precursores que tienen más impacto en cada contaminante, Cabe mencionar que de todos estos contaminantes el único contaminante que tuvo una mayor concentración en 2020 fue el Ozono. Es por eso que durante el reporte se dará un énfasis a los motivos que pudieron dar este fenómeno Para explicar el fenómeno de la concentración del Ozono en 2020 también se buscó como los factores meteorológicos tuvieron un impacto en la alta concentración del ozono en el año. Para esto se utilizó un modelo logístico multinomial.

En conclusión durante este reporte se busca verificar si durante el año de la pandemia 2020 la reducción de actividades tanto civiles como industriales se vio una reducción de los contaminantes en el aire del área Metropolitana de Monterrey, En el caso de que si hubo una menor o mayor concentración en el 2020 se busca respaldar estos resultados con evidencia de informacion historica de las industrias que mas afectan el nivel de concentración de estos contaminantes.

Introducción

En un mundo cada vez más interconectado, la calidad del aire emerge como un tema crucial que demanda nuestra atención colectiva. La contaminación atmosférica, una consecuencia directa de las actividades humanas e industriales, trasciende fronteras y afecta a comunidades enteras, planteando desafíos significativos para la salud pública y la sostenibilidad ambiental. La revisión exhaustiva de la literatura realizada no solo nos proporciona una visión detallada de los contaminantes y sus efectos, sino que también nos insta a considerar la calidad del aire como un barómetro de nuestro compromiso con la salud global y el bienestar de nuestro planeta.

El año 2020, marcado por la inesperada llegada del Covid-19, redefinió nuestra comprensión del mundo y nos obligó a repensar nuestras actividades diarias. Durante este período sin precedentes, las ciudades, una vez bulliciosas y llenas de movimiento, se sumieron en un silencio inusual. Las calles vacías y las fábricas cerradas llevaron a una disminución drástica de las emisiones contaminantes. Este fenómeno, único en la historia moderna, sirvió como un experimento natural que ofrece una oportunidad invaluable para estudiar el impacto directo de nuestras acciones en el medio ambiente. ¿Cómo respondieron los niveles de contaminantes como PM10, NO2, CO y O3 a esta pausa forzada? ¿Qué lecciones podemos extraer de este episodio singular para informar políticas futuras y prácticas industriales más sostenibles?

Los datos recopilados minuciosamente por las estaciones del Sistema Integral de Monitoreo Ambiental (SIMA) se convierten en nuestra brújula en este viaje exploratorio. Al analizar estos datos con precisión y meticulosidad, podemos discernir patrones ocultos y correlaciones significativas. ¿Fueron las concentraciones de contaminantes uniformemente bajas durante el confinamiento? ¿O surgieron disparidades regionales y sectoriales que arrojan luz sobre las áreas de mejora? Este análisis detallado no solo nos permite entender el estado actual de la calidad del aire, sino que también nos empodera para anticipar y mitigar los desafíos futuros.

Al mirar hacia el futuro, es imperativo abordar las complejas intersecciones entre la actividad humana y la salud ambiental. ¿Qué medidas pueden tomar los gobiernos, las industrias y los ciudadanos para garantizar que el regreso a una rutina normal no implique un regreso a niveles insostenibles de contaminación? Este estudio, más allá de su valor intrínseco como exploración científica, se convierte en un llamado a la acción. Nos recuerda nuestra responsabilidad colectiva de preservar la calidad del aire para las generaciones venideras y nos desafía a adoptar prácticas más responsables y sostenibles en nuestra vida diaria.

Para llevar a cabo este estudio, se han empleado datos recopilados durante los años 2020 y 2021 por tres estaciones del Sistema Integral de Monitoreo Ambiental (SIMA). En particular, se han evaluado cuatro contaminantes clave: PM10, NO2, CO y O3. El análisis se centra en comparar las concentraciones de estos contaminantes en las tres estaciones seleccionadas durante ambos años, con el objetivo de determinar si hubo variaciones significativas en los niveles. En caso de encontrar disparidades, se buscará comprender la magnitud de estas diferencias y analizar los posibles factores que pudieron haber contribuido a estos cambios. Además, se investigará de manera detallada las acciones tomadas tanto por la sociedad civil como por la industria para evaluar su impacto en las concentraciones de contaminantes en cada año específico.

Este estudio, respaldado por datos concretos y un enfoque analítico riguroso, aspira a proporcionar una comprensión profunda de cómo las dinámicas sociales y económicas influyen en la calidad del aire que respiramos. Al entender estas complejas interrelaciones, estamos un paso más

cerca de formular políticas y prácticas que preserven tanto la salud humana como la integridad de nuestro medio ambiente para las generaciones futuras.

Justificación de los objetivos

Comprender si hubo diferencias significativas en la calidad del aire entre 2020 y 2021 puede ayudar a evaluar el impacto de la pandemia y las políticas ambientales en la región. Además, identificar las condiciones meteorológicas y las fuentes de emisión relevantes es esencial para tomar medidas efectivas para mejorar la calidad del aire en el futuro.

También durante este estudio se busca ver si las actividades civiles por mas pequeñas que sean pueden tener el mismo impacto que una industria. Durante este estudio se busca encontrar que actividades de nuestro día a día están impactando la concentración de los contaminantes en el aire. Los resultados de este estudio busca concientizar como no solo la industria y las grandes operaciones empresariales pueden perjudicar la calidad del aire sino que todo ciudadano debe poner de su parte para reducir su huella de contaminación.

En muchos casos no es conocimiento público que ciertas prácticas de limpieza o de actividades pueden causar mucha contaminación y durante este reporte se busca encontrar que actividades causan este impacto para educar a los ciudadanos que productos y prácticas se deben limitar.

Descripción del problema específico

El problema específico que abordaremos se refiere a la calidad del aire en Monterrey durante los años 2020 y 2021, con un enfoque en comprender si hubo una diferencia significativa en la calidad del aire entre estos dos años, particularmente debido a la pandemia de COVID-19 que afectó a 2020. Para resolver este problema, se plantean las siguientes preguntas de investigación:

- ¿Hubo una variación significativa en la concentración de los contaminantes clave en el aire de Monterrey entre 2020 y 2021?
- ¿Cuáles fueron las condiciones meteorológicas y las fuentes de emisión relevantes en ambos años y cómo pueden haber influido en la calidad del aire?

Objetivos

- Evaluar y comparar las concentraciones de contaminantes clave en el aire de Monterrey durante los años 2020 y 2021.
- Analizar las condiciones meteorológicas y las fuentes de emisión en ambos años para comprender su influencia en la calidad del aire.
- Identificar cualquier cambio significativo en la calidad del aire que pueda estar relacionado con la pandemia de COVID-19 y otras variables ambientales.

Preparación de los datos

Las estaciones escogidas fueron la Estación Centro, la Estación Noreste², y la Estación Sureste³. También durante este análisis solo se tomarán en cuenta los contaminantes que se incluyen en el Índice de Aire y Salud y las variables meteorológicas proporcionadas por la estación de recolección. Para esta problemática utilizaremos un dataset con 14 variables y 52,462 registros. Por lo que el tamaño del dataset es de 52,462 x 14. Se removieron las variables PM_{2.5}, SO₂, y RAINF dado a que en todas las columnas el número de datos faltantes supera a ocho mil. Para nuestra solución utilizaremos las siguientes variables (Anexo 1). Finalmente durante este Análisis se implementarán dos dataframes uno con las concentraciones numéricas de los contaminantes el cual será usado para las pruebas de Intervalos de Confianza de Medias y la prueba de Anova. Después se generará un dataset con variables categóricas siguiendo la normativa del Socio Formador para clasificar una concentración según el Índice de Calidad de Aire y Salud.

El primer paso para la preparación de los datos fue rellenar el mayor número de datos faltantes posibles. Los datos faltantes para cada columna se calcularon de la siguiente manera. En el caso de que se encontrara un dato faltante en la posición x_i se reemplazará por el promedio del valor anterior y el valor siguiente:

$$x_i = \frac{x_{i-1} + x_{i+1}}{2}$$

También al verificar los datos y sus distribuciones podemos concluir que ninguno sigue una distribución normal.

Al rellenar el mayor número de datos faltantes del dataset numérico podremos generar el dataset categórico. El siguiente paso para transformar el dataset fue estandarizar los valores de los contaminantes según su clasificación en el Índice de Aire y Salud (Anexo 2).

Para que esto pueda ser representada numéricamente las Calificaciones serán representadas de la siguiente manera:

- Buena = 1
- Aceptable = 2
- Malo = 3
- Muy Malo = 4
- Extremadamente Malo = 5

Para lograr estos resultados el primer paso fue convertir los valores de la columna O₃, SO₂ y NO₂ de ppb a ppm ya que esa es la unidad utilizada en el Índice de Aire y Salud. Después se calculó el promedio móvil ponderado según la Normativa de la Secretaría de Gobernación de México. Cada contaminante tiene su tratamiento específico en el cual se considera un número de horas previas diferente para el cálculo de su promedio móvil ponderado. El único Contaminante que no recibe ninguna transformación es CO. En el **Anexo 3** se presentan las transformaciones necesarias para cada variable.

La metodología implementada para el promedio ponderado móvil se puede ver en la siguiente Imagen 1:

$$\bar{C} = \frac{\sum_{i=1}^N C_i W^{i-1}}{\sum_{i=1}^N W^{i-1}}$$

Donde:

$$W = \begin{cases} W & \text{si } W > 0.5 \\ 0.5 & \text{si } W \leq 0.5 \end{cases} \quad y \quad W = 1 - \frac{C_{max} - C_i}{C_{max}}$$

$$\bar{C} = \frac{\sum_{i=1}^N (C_i W^{i-1})}{\sum_{i=1}^N (W^{i-1})}$$

\bar{C} = Concentración promedio móvil ponderada.
 $N = 12$
 Σ = Sumatoria de datos.
 C_i = Concentración promedio horaria de la hora i .
 i = hora consecutiva de medición (la hora más reciente de medición es la hora 1 y la primera hora de medición en el conjunto de datos considerados en el cálculo sería la hora 12).
 W = Factor de ponderación.
 W = Valor del peso.
 C_{max} = Concentración promedio horaria máxima en el periodo de 12 horas.
 C_{min} = Concentración promedio horaria mínima en el periodo de 12 horas.

Imagen 1 : Normativa de Promedio Ponderado Movil

Al tener los criterios necesarios para cada variable se clasificaron según los criterios mencionados en el **Anexo 2**. El último paso en la generación del dataset es crear nuestra variable objetivo. La cual se generó tomando el valor más grande de los contaminantes para cada fila, ya que esta es la normativa implementada por Sima para asignar la calidad de aire que medio la estación en una hora. En este momento ya están terminados los dos datasets que se implementaran para los diferentes análisis. Podrá ver un ejemplo de cómo se ven estos datasets en el (Anexo 4).

Modelación y Validación

Anova Técnica no exitosa

Uno de los Objetivos de este Estudio es verificar si las medias de las concentraciones de los diferentes contaminantes eran iguales en 2020 y 2021. Para esto buscamos implementar un Anova Multifactorial donde el primer factor era el año de la muestra tomada y el segundo factor era la estación donde se tomó la medición. Esta técnica no fue exitosa ya que uno de los supuestos para poder hacer esta prueba es que la población numérica que ingrese al modelo se debe distribuir normalmente. Como se puede ver en el **Anexo 5** ninguna variable se distribuye normalmente en nuestra base de datos. También desarrollamos un modelo de anova y verificamos si este pasaba los supuestos de Normalidad de los Residuos, Independencia, y Homocedasticidad para el modelo. El modelo no obtuvo una calificación satisfactoria en ninguna de estas pruebas como se puede ver en el (Anexo 6). Como el modelo no cumplió ninguno de los supuestos requeridos para implementar un modelo correcto optamos por dejar afuera esta técnica estadística para buscar respuestas para nuestras preguntas objetivo.

Prueba de Intervalos Confianza de Medias con 95% de Confianza

En el siguiente Análisis haremos un Análisis de Intervalos de Confianza de Medias por contaminante. Este análisis se hará comparando las medias de la población de las muestras del año 2020 y 2021 y comparando las medias por estación de sus muestras en 2020 y 2021. La metodología del Análisis es de la siguiente manera. Se calculará un Límite Superior y un Límite inferior y si dentro de este límite se encuentra el cero podremos concluir que no hay diferencias significativas entre la media de los dos poblaciones. En el caso de que el cero no se encuentre dentro del intervalo podremos concluir que si hay diferencias significativas entre las poblaciones. Si el Límite superior es menor a cero la media de la población 2 es mayor a la media de la población 1 y si el Límite superior es mayor a 0 la media de la población 1 es mayor a la media de la población

2. Para esta prueba se cumple el único supuesto que el tamaño de las muestras debe ser mayor a 30. La metodología implementada es la siguiente:

X_1 : Muestras tomadas en 2020 X_2 : Muestras tomadas en 2021

Pruebas de Hipótesis

$$H_0: \mu_1 - \mu_2 \neq 0 \quad H_1: \mu_1 - \mu_2 \neq 0$$

$$\alpha = 0.05 \quad z = 1.96 \text{ o el valor } z \text{ para } \left(\frac{\alpha}{2}\right) \text{ de una dist normal}$$

S_n = Desviación Estándar de la Muestra n_n : Tamaño muestra

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$Linf = (\mu_1 - \mu_2) - (z * SE) \quad Lsup = (\mu_1 - \mu_2) + (z * SE)$$

Se hará esta prueba para cada contaminante y se comparará las mediciones en un año global, y la diferencia entre las mediciones de 2020 y 2021 por estación.

Modelo de Regresión Logística

El siguiente análisis que realizamos es un modelo de regresión logística multinomial, orientado a saber que variables meteorológicas son las más significativas para el índice de calidad del aire del Ozono. Dado que nuestra variable dependiente (índice de calidad del aire del ozono) puede tomar tres valores distintos (1 = Bueno, 2 = Regular, 3 = Malo), un modelo de regresión logística binaria no sería adecuado, por lo que se optó por el enfoque multinomial. A través de este modelo, se evaluaron las variables meteorológicas que estaban en las bases de datos que nos proporcionó la organización socio formadora (SIMA), con el objetivo de identificar su influencia en la calidad del aire con un enfoque en particular en el Ozono. Los resultados del modelo permitieron identificar las variables meteorológicas más significativas y proporcionar una comprensión más profunda de cómo estas condiciones pueden impactar en el índice de calidad del aire del ozono.

El siguiente análisis que realizamos es un estudio de pruebas de hipótesis utilizando el test de Wald para determinar el efecto de una variable independiente sobre la variable dependiente. Las hipótesis que planteamos son las siguientes:

H_0 : La variable independiente no tiene efecto sobre la variable dependiente.

H_1 : La variable independiente tiene un efecto significativo sobre la variable dependiente.

El test de Wald nos permite determinar si un coeficiente estimado en un modelo de regresión es significativamente diferente de cero. Si el resultado del test es significativo, rechazamos la hipótesis nula y concluimos que la variable en cuestión tiene un efecto sobre la variable dependiente. Si el resultado no es significativo, no podemos rechazar la hipótesis nula, sugiriendo que no hay evidencia suficiente para afirmar que la variable tiene un efecto.

Para este análisis, se considera un nivel de significancia típico, como $\alpha = 0.05$. Si el p-valor obtenido del test de Wald es menor que α , rechazamos la hipótesis nula. En cambio, si el p-valor es mayor, no rechazamos la hipótesis nula.

Resultados

Resultados Prueba de Intervalos Confianza de Medias con 95% de Confianza

En el siguiente apartado se presentarán los resultados de la prueba de Intervalos de Confianza de media con 95% de confianza por Contaminante. Y se presentarán las causas externas que respalden estos resultados.

Resultados para Contaminante CO

En la siguiente Tabla 1 se pueden ver los resultados de la prueba de intervalos de confianza de media para CO :

Tabla 1: Resultados Prueba de Intervalos de Confianza de Medias con 95% de Confianza para CO

	Límite Inferior	Límite Superior	Concentración fue mayor en el año:
Año	-0.134	-0.096	2021
Estación Centro	0.173	0.239	2020
Estación Noreste 2	-0.438	-0.36	2021
Estación Sureste3	-0.26	-0.224	2021

En la tabla anterior se puede apreciar como la concentración para CO fue mayor en 2021 globalmente y también fue mayor en 2021 para dos de las tres estaciones evaluadas. En el **Anexo 7** podrán ver las gráficas de los resultados de estas pruebas individuales.

En busca de los motivos externos que causaron esta mayor concentración de CO en 2021 llegamos a la conclusión que el CO se produce en el momento de la combustión de combustibles fósiles para la creación de energía. Según la revista WRI México la actividad que más aporta a la creación de CO es el tráfico vehicular. (*Cuatro Gráficos Que Explican Las Emisiones De Gases De Efecto Invernadero Por País Y Por Sector*, 2021). Durante el año 2021 hubo un incremento drástico en el tráfico vehicular en Monterrey ya que la mayoría de personas empezaban a retomar su rutina y utilizar más los medios de transporte por lo que incrementó la creación de CO en la atmósfera. También la revista PRTR comenta que las industrias que más CO generan son las Industrias Petroleras y las Industrias Metálicas (*CO (Monóxido De Carbono)*, n.d.). Según la base de datos del gobierno de Nuevo León en su base de datos N.L. EMIM Valor Real de Producción Manufacturera en Miles de Pesos (*DATA NUEVO LEÓN | N.L. EMIM Valor Real De Producción Manufacturera*, n.d.). En el **Anexo 8** se puede ver el incremento en Producción de las industrias de metal y de derivados de petróleo en 2021 en comparación a 2020 por lo que tuvo un impacto en su mayor concentración.

Resultados para Contaminante NO2

En la siguiente Tabla 2 se pueden ver los resultados de la prueba de intervalos de confianza de media para NO2:

Tabla 2: Resultados Prueba de Intervalos de Confianza de Medias con 95% de Confianza para NO2

	Límite Inferior	Límite Superior	Concentración fue mayor en el año:
Año	-0.00597	-0.00562	2021
Estación Centro	-0.0083	-0.0076	2021
Estación Noreste 2	-0.001159	-0.01105	2021
Estación Sureste3	-0.00005	0.0001	Concentraciones fueron iguales

En la tabla anterior se puede apreciar como la concentración para NO2 fue mayor en 2021 globalmente y también fue mayor en 2021 para dos de las tres estaciones evaluadas. En el **Anexo 9** podrán ver las gráficas de los resultados de estas pruebas individuales.

En busca de los motivos externos que causaron esta mayor concentración de NO2 en 2021 llegamos a la conclusión que el NO2 también se produce en el momento de la combustión de combustibles fósiles para la creación de energía. Según la revista WRI México la actividad que más aporta a la creación de NO2 es el tráfico vehicular. (*Cuatro Gráficos Que Explican Las Emisiones De Gases De Efecto Invernadero Por País Y Por Sector*, 2021). Durante el año 2021 hubo un incremento drástico en el tráfico vehicular en Monterrey ya que la mayoría de personas empezaban a retomar su rutina y utilizar más los medios de transporte por lo que incrementó la creación de NO2 en la atmósfera. Al igual que el CO la concentración de NO2 se ve más afectada por la industria de metales y de combustibles fósiles como la del petróleo. Según la base de datos del gobierno de Nuevo León en su base de datos N.L. EMIM Valor Real de Producción Manufacturera en Miles de Pesos (*DATA NUEVO LEÓN | N.L. EMIM Valor Real De Producción Manufacturera*, n.d.). En el **Anexo 7** se puede ver el incremento en Producción de la industria de metales y de derivados de petróleo en 2021 en comparación a 2020 por lo que tuvo un impacto en su mayor concentración.

Resultados para Contaminante PM10

En la siguiente Tabla 3 se pueden ver los resultados de la prueba de intervalos de confianza de media para PM10:

Tabla 3: Resultados Prueba de Intervalos de Confianza de Medias con 95% de Confianza para PM10

	Límite Inferior	Límite Superior	Concentración fue mayor en el año:
Año	-5.189	-4.044	2021
Estación Centro	-11.572	-9.707	2021
Estación Noreste 2	-2.951	-1.109	2021
Estación Sureste3	-2.486	-0.294	2021

En la tabla anterior se puede apreciar como la concentración para PM10 fue mayor en 2021 globalmente y también fue mayor en 2021 para las tres estaciones evaluadas. En el **Anexo 10** podrán ver las gráficas de los resultados de estas pruebas individuales.

En busca de los motivos externos que causaron esta mayor concentración de PM10 en 2021 llegamos a la conclusión que el PM10 Las partículas de PM10 se generan en gran medida en la industria por partículas generadas por la construcción y por la minería. Según la Base de datos del gobierno de Nuevo León titulada N.L. Actividad Industrial por Actividad Económica- Variación Anual en el **Anexo 11 (DATA NUEVO LEÓN | N.L. Actividad Industrial Por Actividad Económica- Variación Anual, n.d.)**. Se aprecia un incremento porcentual drástico en la industria de Construcción y Minería en 2021 a comparación de 2021 lo cual genera una concentración de PM10 más alta en 2021.

Resultados para Contaminante O3

En la siguiente Tabla 4 se pueden ver los resultados de la prueba de intervalos de confianza de media para O3

Tabla 4: Resultados Prueba de Intervalos de Confianza de Medias con 95% de Confianza para O3

	Límite Inferior	Límite Superior	Concentración fue mayor en el año:
Año	0.002	0.003	2020
Estación Centro	0.004	0.005	2020
Estación Noreste 2	0.005	0.006	2020
Estación Sureste3	-0.003	-0.002	2021

En la tabla anterior se puede apreciar como la concentración para O3 fue mayor en 2020 globalmente y también fue mayor en 2020 para dos de las tres estaciones evaluadas. En el **Anexo 12** podrán ver las gráficas de los resultados de estas pruebas individuales.

En busca de los motivos externos que causaron esta mayor concentración de PM10 en 2021 llegamos a la conclusión que el ozono se “se forma a partir de reacciones fotoquímicas complejas con intensa luz solar entre contaminantes primarios como son los óxidos de nitrógeno (NO, NO2) y compuestos orgánicos volátiles (COV) “ (*Ozono Troposférico • Ecologistas En Acción*, 2013). Los compuestos orgánicos volátiles

tienen un gran impacto en la concentración de ozono en el aire. Según el estudio Compuestos Orgánicos Volátiles en Área Metropolitana de Monterrey conducido por el INECC se buscó medir los precursores del ozono que estaban más presentes en tres diferentes estaciones de SIMA en el 2015. En **Anexo 13** se podrán ver los 10 precursores que más estuvo presente en cada estación. Según el INECC el precursor del ozono que más estuvo presente en estas estaciones fue el ethanol (Magaña, 2015, 53). Durante la pandemia del covid 19 en el año 2020 se le fue instruido a la población que debía desinfectar todo objeto, ropa, alimento que estuviera en contacto con el exterior. Esto causó que toda la población hiciera un uso excesivo de desinfectantes en aerosol. Estos desinfectantes en aerosol están compuestos de entre 40%-60% de etanol (Anexo 14).

Antes de la pandemia del covid 19 el etanol ya era el precursor del ozono que más estaba presente en el área metropolitana de Monterrey. Durante la pandemia el uso de desinfectantes en aerosol hechos en su mayoría por etanol causó que en 2020 hubiera una mayor concentración de Ozono. A pesar de haber movilidad limitada este fue el único contaminante que tuvo una mayor concentración en 2020.

Resultados Modelo de Regresión Logística

En nuestro modelo de regresión logística, presentamos los coeficientes de las variables independientes, que reflejan su influencia en la variable dependiente. Mediante la prueba de Wald, determinamos la significancia de estas variables. A continuación, detallamos los coeficientes estimados, su interpretación y relevancia estadística.

Se puede observar en el **anexo 15** los coeficientes que nos proporciona el modelo de regresión logística multinomial para el año 2020, donde:

- Intercepto (2 y 3): Estos coeficientes representan el logaritmo de la probabilidad del evento de interés cuando todas las variables independientes son cero. La interpretación específica depende de la naturaleza exacta de la variable dependiente y las categorías en cuestión.
- PRS: Un incremento unitario en PRS se traduce en cambios de -0.0115 y -0.0378 en el logaritmo de las probabilidades de las categorías 2 y 3, respectivamente, manteniendo constantes las demás variables.
- RH: Para cada unidad adicional en RH, los logaritmos de las probabilidades cambiarían en -0.0447 y -0.0554 para las categorías 2 y 3, respectivamente.
- SR: Un incremento unitario en SR se relaciona con un cambio de -0.1986 y -1.7216 en el logaritmo de las probabilidades para las categorías 2 y 3, respectivamente.
- TOUT: Con un aumento de una unidad en TOUT, se espera un cambio en los coeficientes de 0.1088 y 0.2670 en el logaritmo de las probabilidades para las categorías 2 y 3, respectivamente.
- WSR: Por cada incremento unitario en WSR, se prevé una variación de 0.0970 y 0.0317 en el logaritmo de las probabilidades de las categorías 2 y 3.
- WDR: Finalmente, un aumento en WDR resulta en cambios de -0.0047 y -0.0107 en el logaritmo de las probabilidades para las categorías 2 y 3, respectivamente.

Se puede observar en el **anexo 16** los coeficientes que nos proporciona el modelo de regresión logística multinomial para el año 2021, donde:

- Intercepto (2 y 3): Estos valores representan el logaritmo de la probabilidad del evento de interés cuando todas las variables independientes son cero. Los valores difieren entre las categorías, y la interpretación depende de la naturaleza de la variable dependiente.
- PRS: Por cada unidad adicional en PRS, se espera un cambio logarítmico de 0.0061 y -0.0248 en la probabilidad del evento de interés para las categorías 2 y 3, respectivamente, manteniendo constantes las demás variables.

- RH: Para un aumento unitario en RH, los logaritmos de las probabilidades cambiarían en -0.0482 y -0.0249 para las categorías 2 y 3, respectivamente.
- SR: Un incremento unitario en SR se asocia con un cambio de 2.9044 y 3.6864 en el logaritmo de las probabilidades para las categorías 2 y 3, respectivamente.
- TOUT: Con cada unidad adicional en TOUT, los coeficientes indican un cambio de 0.0700 y 0.2321 en el logaritmo de las probabilidades para las categorías 2 y 3.
- WSR: Un aumento en WSR se relaciona con cambios de 0.0146 y -0.0472 en el logaritmo de las probabilidades de las categorías 2 y 3, respectivamente.
- WDR: Por último, para un incremento unitario en WDR, se espera una variación de -0.0048 y -0.0011 en el logaritmo de las probabilidades de las categorías 2 y 3, respectivamente.

Estos coeficientes nos proporcionan una visión de cómo cada variable influye en la probabilidad de que se dé el evento de interés, en función de las categorías 2 y 3 de la variable dependiente.

En el **anexo 17** podemos observar los resultados de la prueba de Wald donde pudimos llegar a la conclusión de que para el año 2020:

Obtuvimos que todas las variables son significativas siendo humedad relativa y temperatura las más significativas para el índice de calidad del aire del ozono.

Para el año 2021 como se puede observar en el **anexo 18** obtuvimos:

que todas las variables son significativas a excepción de la velocidad del viento, siendo radiación solar y humedad relativa las más significativas para el índice de calidad del aire del ozono.

Conclusiones

En conclusión durante esta investigación se utilizaron diferentes herramientas estadísticas para identificar en el año 2021 hubo una mayor concentración en tres de los cuatro contaminantes evaluados. Por ende hubo una peor calidad de aire en 2021 que en 2020.

Para los contaminantes que tuvieron una mayor concentración 2021 que fueron los contaminantes PM10, CO, y NO2 su incremento en la concentración se debió en gran medida por la reactivación de la economía y la industria en el año 2021 al salir de la pandemia de Covid. Es un claro ejemplo de cómo es importante regular las diferentes industrias y cómo diferentes procesos industriales afectan la concentración de diferentes contaminantes. Como por ejemplo con el contaminante PM10 donde gracias a la base de datos del gobierno de Nuevo León se demostró que hubo un incremento en la actividad económica de los precursores que más tienen efecto en la concentración de PM10. Para el caso de CO y NO2 los precursores que causaron una mayor concentración en 2021 fueron el regreso a la normalidad de la rutina de las personas en 2021. Lo que causó un incremento en los autos en las calles de Monterrey, la industria del metal y la industria del petróleo y sus precursores. Mediante la investigación de cómo estos precursores se comportaron tanto en 2020 y 2021 podemos respaldar nuestros resultados estadísticos y dar una idea más clara de porque recibimos esos resultados.

Por otro lado, el contaminante O3 fue el único que tuvo una mayor concentración en 2020. Este es un claro ejemplo de cómo no sólo la industria y las diversas actividades económicas tienen un impacto en la calidad del aire. Para este caso vimos como una instrucción de las autoridades sanitarias para mitigar la pandemia del covid 19 como lo fue desinfectar cualquier objeto que haya tenido contacto con el exterior con desinfectantes en aerosol al ser ejecutada por toda la población por un tiempo prolongado de tiempo. Causó una mayor concentración de O3 en el año 2020.

Durante el año 2020, nuestra investigación reveló que todas las variables meteorológicas en estudio fueron significativas para determinar el índice de calidad del aire del ozono. Específicamente, la humedad relativa y la temperatura sobresalieron como las más influyentes. Por otro lado, en el 2021, si bien la mayoría de las variables mantuvieron su significancia, la velocidad del viento no tuvo un impacto notable. En este año, la radiación solar y la humedad relativa emergieron como las variables más determinantes para el índice de calidad del aire del ozono.

En conclusión la calidad del aire en 2021 en efecto fue peor que en 2020. Tres contaminantes tuvieron una mayor concentración en 2021 que en 2020. Finalmente, en este estudio se demostró que no solo la industria tiene un efecto negativo en la calidad del aire de la ciudad, sino que acciones pequeñas como desinfectar objetos al producirse a gran escala pueden llegar a tener un efecto igual de negativo que toda una industria la hora de afectar la calidad de aire de una área metropolitana. Para poder mitigar las altas concentraciones de estos contaminantes se debe hacer un trabajo conjunto y cooperativa entre la empresa privada, las diferentes entidades gubernamentales, y la población como un todo para detectar y reducir las actividades que tienen más impacto en el comportamiento de estos contaminantes.

Liga de Google Drive a Folder con códigos en Python y R del Trabajo

<https://drive.google.com/drive/folders/1fuYWXOz1Dj7pQSrnmh4mIJmnz0Udlts?usp=sharing>

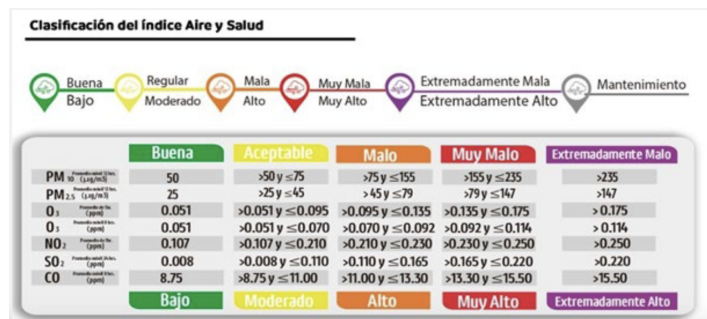
Anexos

Anexo 1 Lista de Variables Utilizadas

- date: Variable Categórica
- Año: Variable Categórica
- Estacion: Variable Categórica ○ Esta variable representa la estación donde fue recolectada la información.
- CO: Variable Numérica ○ Esta variable representa el nivel del contaminante CO en el índice de calidad del aire y salud medido por el promedio móvil ponderado medido en la estación durante una hora. Es importante porque queremos descubrir el impacto de este contaminante en la calidad del aire.
- NO2: Variable Numérica ○ Esta variable representa el nivel del contaminante NO2 en el índice de calidad del aire y salud medido por el promedio móvil ponderado medido en la estación durante una hora. Es importante porque queremos descubrir el impacto de este contaminante en la calidad del aire.
- O3: Variable Numérica ○ Esta variable representa el nivel del contaminante O3 en el índice de calidad del aire y salud medido por el promedio móvil ponderado medido en la estación durante una hora. Es importante porque queremos descubrir el impacto de este contaminante en la calidad del aire.
- PM10: Variable Numérica ○ Esta variable representa el nivel del contaminante PM10 en el índice de calidad del aire y salud medido por el promedio móvil ponderado medido en la estación durante una hora. Es importante porque queremos descubrir el impacto de este contaminante en la calidad del aire.
- O3: Variable Numérica ○ Esta variable representa el nivel del contaminante NO2 en el índice de calidad del aire y salud medido por la concentración medida en la estación durante una hora. Es importante porque queremos descubrir el impacto de este contaminante en la calidad del aire.
- PRS: Variable Numérica ○ Esta variable representa la presión atmosférica en la estación durante una hora. Es importante para el modelo ya que queremos descubrir que tanto este factor meteorológico impactó con la calidad del aire.
- RH: Variable Numérica ○ Esta variable representa el porcentaje de humedad relativa en la estación durante una hora. Es importante para el modelo ya que queremos descubrir que tanto este factor meteorológico impactó con la calidad del aire.
- SR: Variable Numérica ○ Esta variable representa la radiación solar en la estación durante una hora. Es importante para el modelo ya que queremos descubrir que tanto este factor meteorológico impactó con la calidad del aire.

- TOUT: Variable Numérica ○ Esta variable representa la temperatura promedio en la estación durante una hora. Es importante para el modelo ya que queremos descubrir que tanto este factor meteorológico impactó con la calidad del aire.
- WSR: Variable Numérica ○ Esta variable representa la velocidad del viento en la estación durante una hora. Es importante para el modelo ya que queremos descubrir que tanto este factor meteorológico impactó con la calidad del aire.
- WDR: Variable Numérica ○ Esta variable representa la dirección del viento en la estación durante una hora. Es importante para el modelo ya que queremos descubrir que tanto este factor meteorológico impactó con la calidad del aire.
- Objetivo: Variable Numérica ○ El valor del Índice de Calidad de Aire más alto medido cada hora en la estación. Es importante porque es la variable que trataremos de predecir

Anexo 2 Imagen de Normativa de Índice de Aire y Salud



Anexo 3 Transformaciones Necesarias por Contaminante para el Índice de Aire y Salud.

Contaminante	Concentración base
PM10	Concentración promedio móvil ponderado de 12 horas*
PM2.5	
ozono (O3)	Concentración promedio móvil de 8 horas
monóxido de carbono (CO)	
dióxido de nitrógeno (NO2)	Concentración promedio horaria
ozono (O3)	
dióxido de azufre (SO2)	Concentración promedio móvil de 24 horas (como aproximación al promedio de 24 horas)

Anexo 4 Ejemplo de Datasets Utilizados en el Estudio

Dataset Numérico:

date	CO	NO2	PM10	PRS	RH	SR	TOUT	WSR	WDR	O3Concentracion	Estacion	Year
2020-01-01 00:00:00	NA	NA	86.34000	NA	NA	0.000	NA	NA	NA	NA	Centro_2020	2020
2020-01-01 01:00:00	NA	0.0065	112.01000	713.6	91	0.158	10.49	3.7	2	0.019	Centro_2020	2020
2020-01-01 02:00:00	NA	0.0056	100.01000	712.8	91	0.156	10.51	1.9	144	0.018	Centro_2020	2020

Dataset Categórico:

date	CO	NO2	PM10	PRS	RH	SR	TOUT	WSR	WDR	O3Concentracion	Estacion	Year	Objetivo
2020-02-01 00:00:00	1	1	1	708.7	88	0.156	12.30	1.4	263	1	Centro_2020	2020	1
2020-02-01 01:00:00	1	1	1	708.1	85	0.155	12.41	3.1	239	1	Centro_2020	2020	1
2020-02-01 02:00:00	1	1	1	707.2	81	0.155	12.88	3.2	239	1	Centro_2020	2020	1
2020-02-01 03:00:00	1	1	1	707.0	77	0.155	13.22	3.9	258	1	Centro_2020	2020	1

Anexo 5 Prueba de Normalidad a las diferentes variables

H_0 : Los datos provienen de una población que sigue distribución normal

H_1 : Los datos provienen de una población que no sigue distribución normal

$$\alpha = 0.05$$

El desglose de la prueba de normalidad para las variables se puede ver en la siguiente Imagen donde se le asigna un valor de 1 cuando no se Rechaza H_0 y la población sigue una distribución normal. Se asigna un 0 cuando se rechaza H_0 y la población no sigue una distribución normal.

	Dataset	CO	NO2	O3	PM10	PM2.5	PRS	RAINP	RH	SO2	SR	TOUT	WSR	WDR	Total_Estacion
0	Centro_2020	1	0	0	0	0	0	1	0	0	0	0	0	0	2
1	Noreste2_2020	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Sureste3_2020	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Centro_2021	0	0	0	0	0	0	1	0	0	0	0	0	0	1
4	Noreste2_2021	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	Sureste3_2021	1	0	0	0	0	0	0	0	0	0	0	0	0	1

Anexo 6 Prueba de Supuestos para modelo de Anova

Prueba de Normalidad de Residuos Anderson Darling:

$$\alpha: 0.05$$

H_0 : Los residuos se distribuyen normalmente

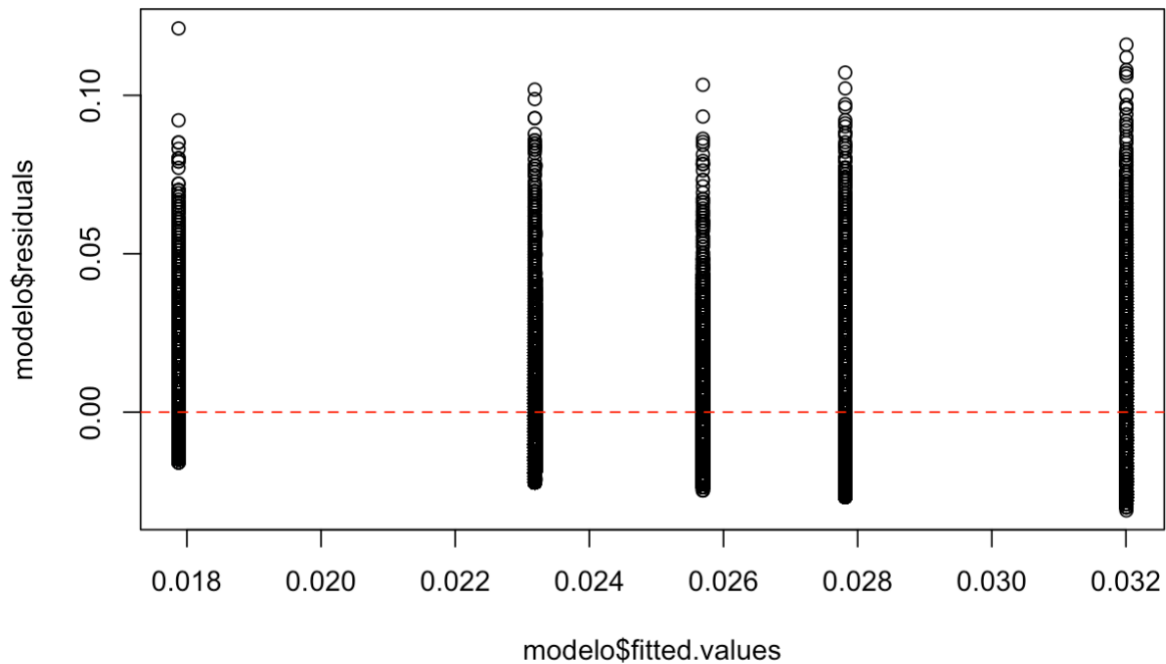
H_1 : Los residuos no se distribuyen normalmente

Anderson-Darling normality test

```
data: modelo$residuals
A = 607.93, p-value < 2.2e-16
```

Como el p valor =0 tengo suficiente información para rechazar H_0 y concluir que los residuos no se distribuyen normalmente.

Gráfica de Homocedasticidad de los residuos.



En esta gráfica se puede ver como en todos los valores de x en el eje de los residuales la gran mayoría de los valores se encuentran en posiciones mayores a 0 por lo que lo que aparece en el eje positivo no se refleja en el eje negativo por lo que este modelo no muestra homocedasticidad de los residuos.

Prueba de Independencia Durbin Watson

$\alpha: 0.05$

$$H_0: Cov(u_i, u_j) = 0$$

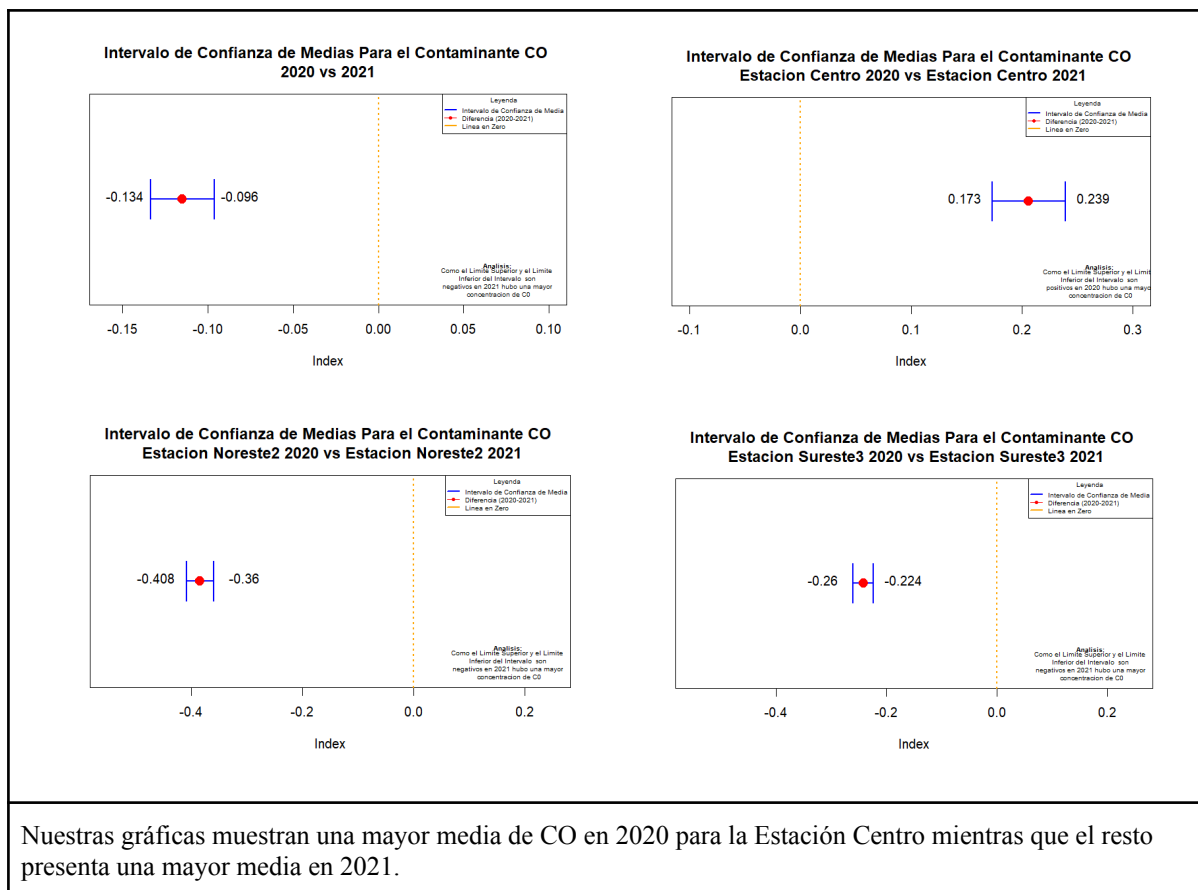
$$H_1: Cov(u_i, u_j) \neq 0$$

Durbin-Watson test

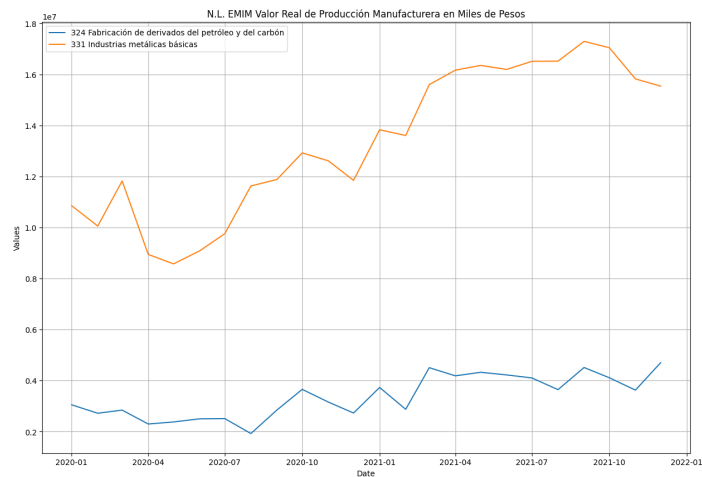
```
data: modelo
DW = 0.15759, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

Como el p valor $\neq 0$ tengo suficiente información para rechazar H_0 y concluir que no son independientes.

Anexo 7 Gráficas de Pruebas de Intervalos de confianza de medias para CO

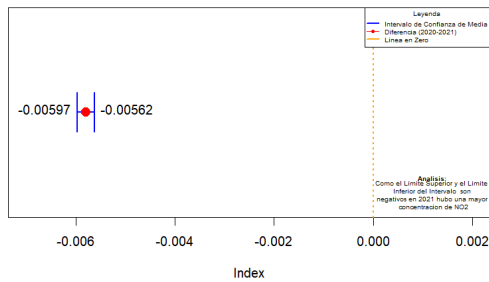


Anexo 8 Gráfica de Actividad en las Industrias metálicas y de derivados de Petróleo proporcionada por el gobierno de Nuevo León para el año 2020 y 2021

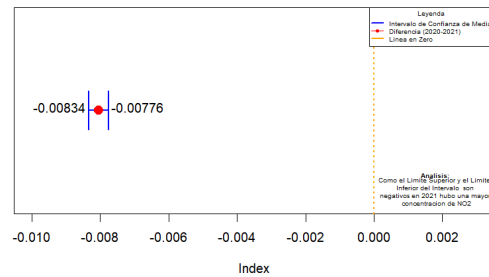


Anexo 9 Gráficas de Pruebas de Intervalos de confianza de medias para NO2

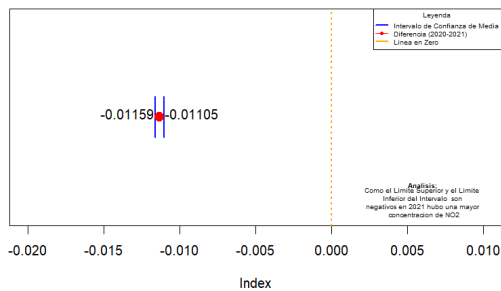
**Intervalo de Confianza de Medias Para el Contaminante NO2
2020 vs 2021**



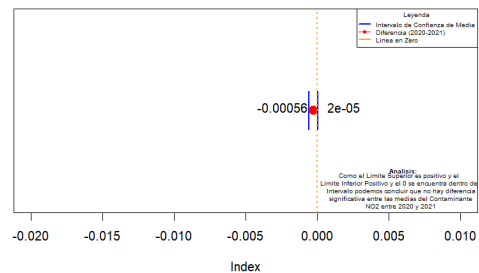
**Intervalo de Confianza de Medias Para el Contaminante NO2
Estacion Centro 2020 vs Estacion Centro 2021**



**Intervalo de Confianza de Medias Para el Contaminante NO2
Estacion Noreste2 2020 vs Estacion Noreste2 2021**



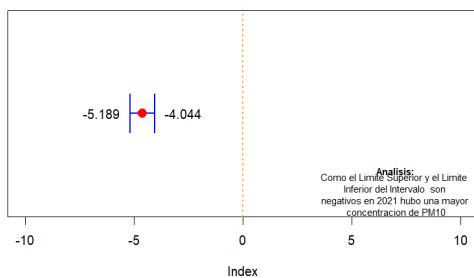
**Intervalo de Confianza de Medias Para el Contaminante NO2
Estacion Sureste3 2020 vs Estacion Sureste3 2021**



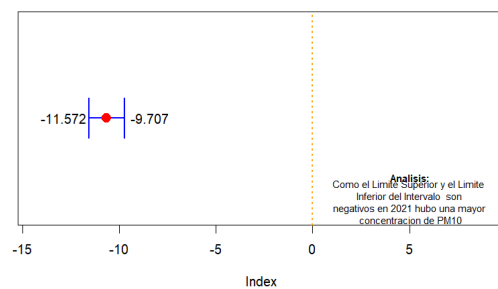
La mayoría de nuestras gráficas indican que en 2021 las medias de NO2 fueron considerablemente mayores, sin embargo en la Estación Sureste las medias de ambos años son bastante similares.

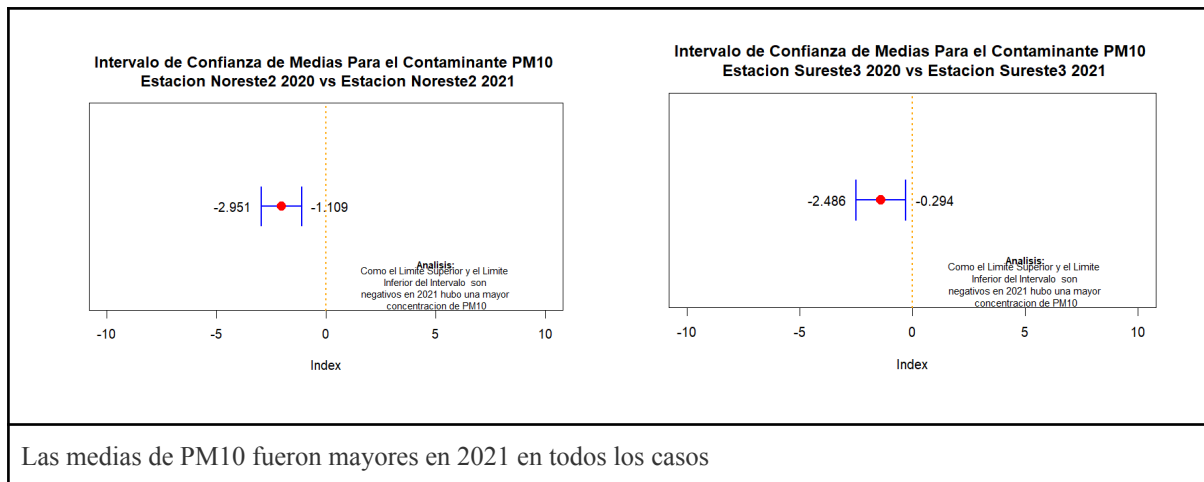
Anexo 10 Gráficas de Pruebas de Intervalos de confianza de medias para PM10

**Intervalo de Confianza de Medias Para el Contaminante PM10
2020 vs 2021**

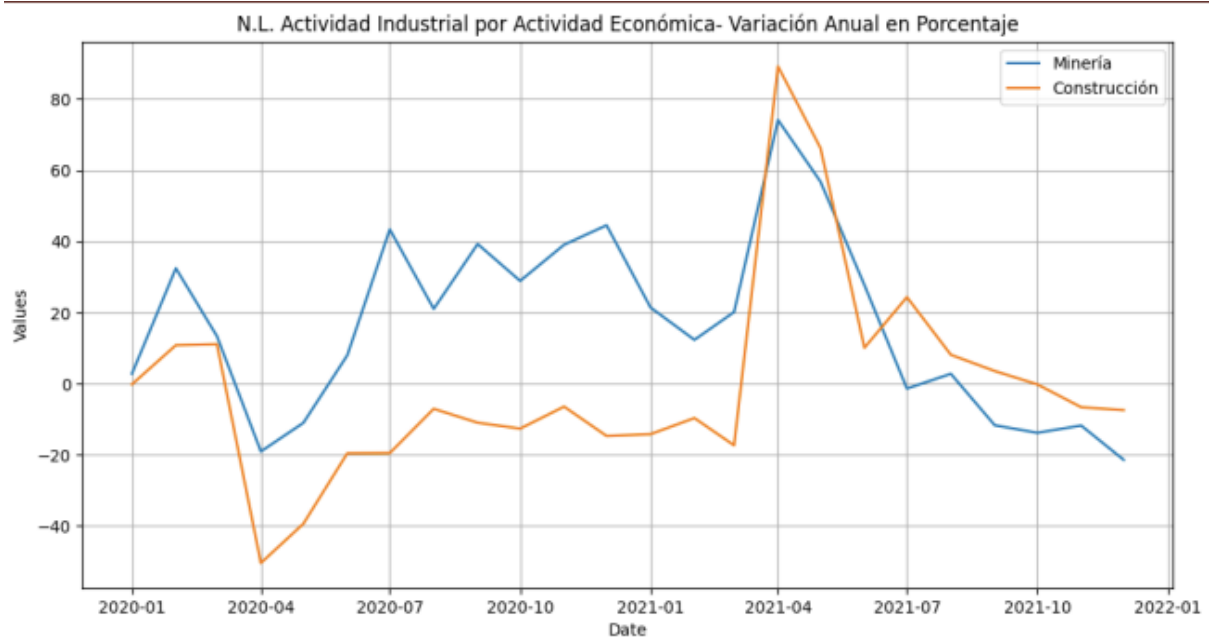


**Intervalo de Confianza de Medias Para el Contaminante PM10
Estacion Centro 2020 vs Estacion Centro 2021**

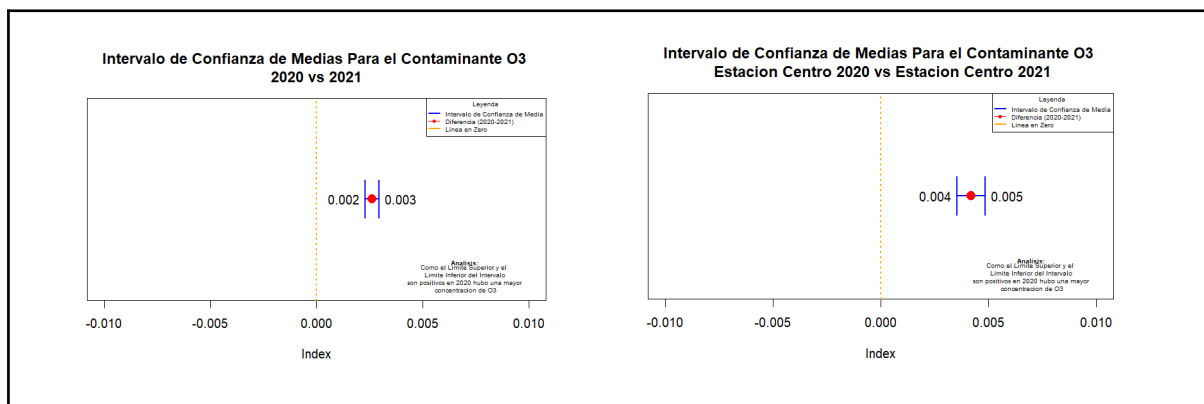


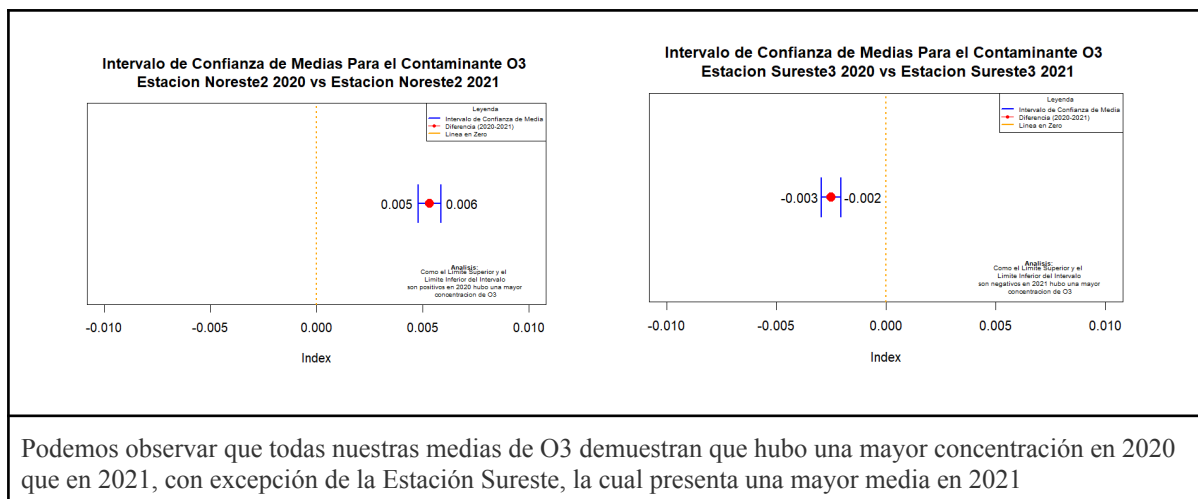


Anexo 11 Gráfica de Actividad en las Industrias de Construcción y Minería proporcionada por el gobierno de Nuevo León para el año 2020 y 2021.



Anexo 12 Gráficas de Pruebas de Intervalos de confianza de medias para O3





Anexo 13 Tabla de Precursores del ozono que estaban más presentes en el área metropolitana de Monterrey.

Tabla 3. Los diez compuestos comunes a los tres sitios que más contribuyen a la FPO total. Se indica la contribución individual de cada compuesto en cada sitio a la FPO.

Santa Catarina	San Nicolás	San Bernabé
etanol (alcohol etílico): 233.5 ppbV O ₃	etanol (alcohol etílico): 52.46 ppbV O ₃	etanol (alcohol etílico): 51.07 ppbV O ₃
tolueno: 101.00 ppbV O ₃	tolueno: 104.16 ppbV O ₃	tolueno: 38.98 ppbV O ₃
propileno: 82.00 ppbV O ₃	propileno: 46.52 ppbV O ₃	propileno: 32.61 ppbV O ₃
1,2,4-trimetilbenceno: 79.52 ppbV O ₃	1,2,4-trimetilbenceno: 87.73 ppbV O ₃	1,2,4-trimetilbenceno: 118.38 ppbV O ₃
o-xileno: 79.44 ppbV O ₃	o-xileno: 53.19 ppbV O ₃	o-xileno: 50.84 ppbV O ₃
m-xileno: 78.37 ppbV O ₃	m-xileno: 52.46 ppbV O ₃	m-xileno: 62.27 ppbV O ₃
p-xileno: 71.83 ppbV O ₃	p-xileno: 55.37 ppbV O ₃	p-xileno: 53.70 ppbV O ₃
4-metil-2-pentanona (metil isobutil cetona): 59.18 ppbV O ₃	4-metil-2-pentanona (metil isobutil cetona): 53.12 ppbV O ₃	4-metil-2-pentanona (metil isobutil cetona): 46.46 ppbV O ₃
1,3,5-trimetilbenceno (mesitileno): 50.94 ppbV O ₃	1,3,5-trimetilbenceno (mesitileno): 65.18 ppbV O ₃	1,3,5-trimetilbenceno (mesitileno): 94.02 ppbV O ₃
n-hexano: 50.28 ppbV O ₃	n-hexano: 78.06 ppbV O ₃	n-hexano: 71.36 ppbV O ₃

Anexo 14 Etiqueta de los compuestos de un desinfectante en aerosol

3. Composición / Información sobre los ingredientes		
Ingrediente(s)	# CAS	Porcentaje
Etanol	64-17-5	40 - 60
Butano	106-97-8	1 - 5
Propano	74-98-6	1 - 5
Alquilo (50% C14, 40% C12, 10% C16) dimetil bencil amonio sacarinato	No se aplica	0 - 0.1

Anexo 15 Coeficientes de regresión logística multinomial para el año 2020

Coefficients:	(Intercept)	PRS	RH	SR	TOUT	WSR	WDR
2	4.792634	-0.01153557	-0.04447838	-0.1986845	0.1088022	0.09700283	-0.004790449
3	17.447508	-0.03738808	-0.05544725	-1.7216879	0.2670784	0.03169910	-0.010733776

Anexo 16 Coeficientes de regresión logística multinomial para el año 2021

(Intercept)	PRS	RH	SR	TOUT	WSR	WDR
2	-7.082130	0.006137756	-0.04822663	2.904467	0.07008665	0.01461195
3	6.148726	-0.024837003	-0.02497572	3.686489	0.23219355	-0.04726438

Anexo 17 Resultados prueba de Wald para el año 2020

```
Analysis of Deviance Table (Type II tests)

Response: O3Concentracion
      LR Chisq Df      Pr(>Chisq)
PRS    16.76  2      0.0002291 ***
RH    510.14  2 < 0.0000000000000022 ***
SR    13.65  2      0.0010889 **
TOUT  493.53  2 < 0.0000000000000022 ***
WSR   169.40  2 < 0.0000000000000022 ***
WDR   138.06  2 < 0.0000000000000022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anexo 18 Resultados prueba de Wald para el año 2021

```
Analysis of Deviance Table (Type II tests)

Response: O3Concentracion
      LR Chisq Df      Pr(>Chisq)
PRS     9.95  2      0.006926 **
RH    538.31  2 < 0.0000000000000022 ***
SR    661.05  2 < 0.0000000000000022 ***
TOUT  157.04  2 < 0.0000000000000022 ***
WSR     5.56  2      0.062006 .
WDR   103.88  2 < 0.0000000000000022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Referencias

- Magaña, M. (2015).). *Evaluación de Compuestos Orgánicos Volátiles en el Área Metropolitana de Monterrey*. Instituto Nacional de Ecología y Cambio Climático (INECC).
https://www.gob.mx/cms/uploads/attachment/file/370440/8_Informe_Final_de_COVs_Monterrey.pdf
- Clarity. (2023, Marzo 21). The purpose and importance of air quality monitoring. Recuperado de <https://www.clarity.io/blog/what-is-air-quality-monitoring-why-is-it-important>
- UNEP. (2021, Noviembre 2). 5 dangerous pollutants you're breathing in every day. Recuperado de <https://www.unep.org/news-and-stories/story/5-dangerous-pollutants-youre-breathing-every-day>
- ay NOAA SciJinks. (2023). ¿Cómo se mide la calidad del aire? Recuperado de <https://scijinks.gov/air-quality/>
- Organización Mundial de la Salud. (2021, Septiembre 22). Nuevas directrices mundiales de la OMS sobre la calidad del aire tienen como objetivo salvar

millones de vidas de la contaminación del aire. Recuperado de

[https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-sa](https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution)

ve-millions-of-lives-from-air-pollution Norma Oficial Mexicana

(2019) https://www.dof.gob.mx/nota_detalle.php?codigo=5579387&fecha=20/11/2019#gsc.tab=0