



## **ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**

**Modelo matemático para relacionar el número de contagios del COVID-19 y los contaminantes ambientales en la ciudad de Quito Ecuador, periodo enero 2020 – marzo 2021**

**CARLOS GABRIEL PÉREZ CASTILLO**

**Trabajo de Titulación modalidad: Proyectos de Investigación y Desarrollo, presentado ante el Instituto de Posgrado y Educación Continua de la ESPOCH, como requisito parcial para la obtención del grado de:**

**MAGÍSTER EN MATEMÁTICA MENCIÓN MODELACIÓN Y DOCENCIA**

**RIOBAMBA – ECUADOR**

**NOVIEMBRE, 2022**

**@2022, Carlos Gabriel Pérez Castillo**

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.



## ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

EL TRIBUNAL DE TRABAJO DE TITULACIÓN CERTIFICA QUE:

El Trabajo de Titulación modalidad Proyectos de Investigación y Desarrollo, titulado **Modelo matemático para relacionar el número de contagios del COVID-19 y los contaminantes ambientales en la ciudad de Quito Ecuador, periodo enero 2020 – marzo 2021**, de responsabilidad del señor Carlos Gabriel Pérez Castillo ha sido prolijamente revisado y se autoriza su presentación.

Dr. Juan Mario Vargas Guambo. Mag.



Firmado electrónicamente por:  
**JUAN MARIO  
VARGAS GUAMBO**

### **PRESIDENTE**

Lic. Andrea Damaris Hernández Allauca, Mag.



Firmado electrónicamente por:  
**ANDREA DAMARIS  
HERNANDEZ  
ALLAUCA**

### **DIRECTORA**

Ing. Deysi Margoth Guanga Chunata, Mag



Firmado electrónicamente por:  
**DEYSI MARGOTH  
GUANGA CHUNATA**

### **MIEMBRO**

Ing. Guicela Margoth Ati Cutiupala, Mag.



Firmado electrónicamente por:  
**GUCELA  
MARGOTH ATI  
CUTIUPALA**

### **MIEMBRO**

Riobamba, noviembre de 2022

## **DERECHOS INTELECTUALES**

Yo, Carlos Gabriel Pérez Castillo, declaro que soy responsable de las ideas, doctrinas y resultados expuestos en el **Trabajo de Titulación modalidad Proyecto de Investigación y Desarrollo**, y que el patrimonio intelectual generado por la misma pertenece exclusivamente a la Escuela Superior Politécnica de Chimborazo.

---

Carlos Gabriel Pérez Castillo

0201871050

## DECLARACIÓN DE AUTENTICIDAD

Yo, Carlos Gabriel Pérez Castillo, declaro que el presente **Trabajo de Titulación modalidad Proyecto de Investigación y Desarrollo**, es de mi auditoria y que los resultados de este son auténticos y originales. Los textos constantes en el documento que provienen de otra fuente están debidamente citados y referenciados.

Como autor, asumo la responsabilidad legal y académica de los contenidos de este proyecto de investigación de maestría.

---

Carlos Gabriel Pérez Castillo

0201871050

## **DEDICATORIA**

A Dios quien ha sido mi guía, fortaleza que sus manos de fidelidad y amor han estado conmigo hasta el día de hoy. A mis padres CARLOS y ANITA quienes con su amor, paciencia y esfuerzo me han permitido llegar a cumplir hoy un sueño más, gracias por inculcar en mí el ejemplo de esfuerzo y valentía, de no temer las adversidades. A mis hermanos por su cariño y apoyo incondicional, durante todo este proceso, por estar conmigo en todo momento.

Con amor dedico esta investigación a mi pequeña hija FRANSHESKA RAPHAELA, quien es mi motivación para seguir alcanzando mis sueños y metas. A mi esposa JANETH ALEXANDRA, quien es mi apoyo y fortaleza en todo momento. A toda mi familia porque con sus oraciones, consejos y palabras de aliento hicieron de mí una mejor persona que de una u otra forma me acompañan.

CARLOS

## **AGRADECIMIENTO**

En primer lugar, mis agradecimientos a la Escuela Superior Politécnica del Chimborazo, por brindarme la oportunidad de adquirir los conocimientos necesarios para afrontar la vida profesional. A las autoridades, personal docente y administrativo de Escuela Superior Politécnica del Chimborazo, por su valioso apoyo en todo el proceso de formación y obtención del Título.

Carlos

## TABLA DE CONTENIDO

RESUMEN .....	xv
SUMMARY / ABSTRACT .....	xvi
CAPÍTULO I.....	1
1. INTRODUCCIÓN .....	1
1.1. Situación problemática .....	1
1.2. Formulación del problema.....	2
1.3. Situación del problema.....	2
1.4. Formulación del problema.....	3
1.5. Preguntas directrices.....	3
1.6. Justificación.....	4
1.7. Objetivos .....	5
1.7.1. Objetivo general .....	5
1.7.2. Objetivos específicos.....	5
1.8. Hipótesis.....	6
1.8.1. Hipótesis general .....	6
1.9. Identificación de variables.....	6
CAPÍTULO II.....	7
2. MARCO TEÓRICO .....	7
2.1. Antecedentes del problema.....	7
2.1.1 Modelos de regresión línea.....	12
2.1.1.1 Regresión lineal simple .....	12
2.1.1.2 Distribución F de Fisher .....	16
2.1.1.3 Inferencia de Regresión lineal para $\beta_0$ y $\beta_1$ .....	16



2.1.1.4	Los intervalos de confianza para cada $\beta_1$ .....	17
2.1.1.5	Regresión lineal múltiple.....	17
2.1.1.	Ajuste por mínimos cuadrados .....	18
2.2.	Desarrollo de modelos matemáticos .....	22
2.3.	Supuestos del Modelo.....	22
2.3.1.	Validación de Modelos Desde la Perspectiva Epistemológica.....	22
2.3.2.	El Modelo.....	24
2.3.3.	El proceso del M&S .....	24
CAPÍTULO III .....		28
3.	METODOLOGÍA DE INVESTIGACIÓN.....	28
3.1	Tipo y diseño de la investigación .....	28
3.2	Métodos de investigación .....	29
3.3	Enfoque de la investigación.....	30
3.4	Alcance de la investigación .....	30
3.5	Recolección de datos para la evaluación del estudio .....	31
3.5.1	Población.....	31
3.6	Datos .....	32
3.6.1	La regresión lineal.....	32
3.6.2	Regresión lineal simple .....	32
3.6.2	Regresión lineal múltiple.....	33
3.6.3	Regresión lineal multivariante.....	34
3.6.4	Regresión lineal múltiple multivariante.....	34
3.7	Selección de Muestra.....	35
3.8	Tamaño de Muestra .....	35
3.9	Correlación de Spearman .....	36
3.10	Pasos para la prueba del coeficiente de correlación de Spearman.....	36
3.10.1	Formulación de hipótesis.....	36
3.10.2	Nivel de significancia “ $\alpha$ ” .....	37

CAPÍTULO IV .....	40
4. RESULTADOS Y DISCUSIÓN .....	40
4.1 Generalidades .....	40
4.1.1 Contagios de Covid 19 .....	40
4.1. Contaminantes atmosféricos .....	42
4.2. Prueba de Normalidad .....	46
4.3. Correlación de Pearson y Spearman .....	48
4.4. Regresión lineal simple .....	51
4.4.1. Modelo 1: Contagios nuevos de Covid vs CO <sub>2</sub> .....	52
4.4.2. Modelo 2: Contagios Covid en el periodo marzo 2020 – marzo 2021 vs O <sub>3</sub> .....	53
4.4.3. Modelo 3: Contagios Covid en el periodo marzo 2020 – marzo 2021 vs SO <sub>2</sub> .....	54
4.4.4. Modelo 4: Contagios Covid en el periodo marzo 2020 – marzo 2021 vs CO .....	55
4.5. Regresión lineal múltiple.....	56
4.5.1. Modelo 1 .....	56
4.5.2. Modelo 2 .....	59
4.8 Decisión de prueba de hipótesis .....	60
4.9 Análisis comparativo en R.....	61
4.9.1 Regresión Lineal .....	61
4.9.1.1 Modelo 1: Contagio Covid vs X1_CO2.....	61
4.9.1.2 Modelo 2: Contagio Covid vs X2_03 .....	63
4.9.1.3 Modelo 3: Contagio Covid vs X2_SO2 .....	64
4.9.1.4 Modelo 4: Contagio Covid vs X2_CO.....	66
4.8.2 Mejor modelo.....	67
4.8.3 Prueba de normalidad.....	68
4.8.3.1 Shapiro Test .....	68
4.8.3.2 Kolmogorov – Smirnov .....	69
4.9 Regresión Lineal Múltiple En R - Studio .....	69
4.9.1 Relación entre variables .....	69
4.9.2 Correlación de Pearson .....	70

CONCLUSIONES.....	74
RECOMENDACIONES.....	75
GLOSARIO	
BIBLIOGRAFÍA	
ANEXOS	

## ÍNDICE DE TABLA

<b>Tabla 1-2</b>	Análisis de Varianza .....	15
<b>Tabla 2-2</b>	Ejemplo de mínimos cuadrados .....	21
<b>Tabla 3-2</b>	Ejemplo de mínimos cuadrados .....	21
<b>Tabla 4-4</b>	Contagios Covid 19 .....	41
<b>Tabla 5-4</b>	Contaminantes .....	42
<b>Tabla 6-4</b>	Valores de concentración de contaminantes atmosféricos mg/m <sup>3</sup> .....	43
<b>Tabla 7-4</b>	Análisis de normalidad de los datos.....	47
<b>Tabla 8-4</b>	Valores de la correlación .....	49
<b>Tabla 9-4</b>	Valores de la correlación de los contaminantes con los casos de Covid-19 .....	50
<b>Tabla 10-4</b>	Valores de R de los modelos matemáticos .....	51
<b>Tabla 11-4</b>	Coefficientes del modelo 1.....	52
<b>Tabla 12-4</b>	Coefficientes del modelo 2.....	53
<b>Tabla 13-4</b>	Coefficientes del modelo 3.....	54
<b>Tabla 14-4</b>	Coefficientes del modelo 4.....	55
<b>Tabla 15-4</b>	Resultados de los análisis de regresión lineal múltiple.....	56
<b>Tabla 16-4</b>	ANOVA aplicado al modelo matemático en SPSS .....	57
<b>Tabla 17-4</b>	Coefficientes beta del modelo matemático 1 .....	57
<b>Tabla 18-4</b>	Resultados de los análisis de regresión lineal múltiple del modelo 2 .....	59
<b>Tabla 19-4</b>	ANOVA aplicado al modelo matemático en SPSS .....	59
<b>Tabla 20-4</b>	Coefficientes beta del modelo matemático 2.....	60
<b>Tabla 21-4</b>	SHAPIRO – WILK TEST DE NORMALIDAD .....	68
<b>Tabla 22-4</b>	Kolmogorov – Smirnov test de normalidad.....	69

## ÍNDICE DE FIGURAS

<b>Figura 1-2.</b>	Métodos mínimos cuadrados .....	19
<b>Figura 2-2.</b>	Ejemplo de mínimos cuadrados.....	22
<b>Figura 3-2.</b>	Proceso de M&S en función de la validez del modelo obtenido .....	25
<b>Figura 4-2.</b>	Ciclo del modelado y simulación.....	27
<b>Figura 5-3.</b>	Ejemplo de regresión lineal simple.....	33
<b>Figura 6-3</b>	Regresión lineal múltiple, millas por galón (MPG) .....	33
<b>Figura 7-3.</b>	Ejemplo de regresión lineal multivariante .....	34
<b>Figura 8-3.</b>	Regresión lineal múltiple multivariante que calcula las MPG .....	35
<b>Figura 9-4.</b>	Comportamiento de contagios de Covid-19 marzo 2020 – marzo 2021.....	41
<b>Figura 11-4.</b>	Comportamiento de los contaminantes atmosféricos .....	44
<b>Figura 12-4.</b>	Relación de los contaminantes atmosféricos y los contagios de Covid-19.....	44
<b>Figura 13-4.</b>	Relación de los contaminantes atmosféricos y los contagios de covid .....	45
<b>Figura 14-4.</b>	Gráficos de pruebas de normalidad.....	48
<b>Figura 15-4.</b>	Ingreso de datos al programa SPSS .....	49
<b>Figura 16-4.</b>	Salida de datos del programa SPSS .....	49
<b>Figura 17-4.</b>	Análisis de correlación en Matlab.....	50
<b>Figura 18-4.</b>	Regresión lineal de CO <sub>2</sub> vs contagios nuevos de Covid .....	52
<b>Figura 20-4.</b>	Regresión lineal de SO <sub>2</sub> vs contagios nuevos de Covid.....	54
<b>Figura 21-4.</b>	Regresión lineal de CO vs contagios nuevos de Covid.....	55
<b>Figura 22-4.</b>	Regresión lineal múltiple del modelo 1 .....	58
<b>Figura 23-4.</b>	Regresión lineal múltiple del modelo 2 .....	60
<b>Figura 25-4.</b>	Regresión lineal simple Contagio Covid vs X <sub>1</sub> _CO <sub>2</sub> .....	61
<b>Figura 27-4.</b>	Regresión lineal simple Casos Covid – X <sub>2</sub> _O <sub>3</sub> .....	63
<b>Figura 29-4.</b>	Regresión lineal simple Casos Covid – X <sub>3</sub> _SO <sub>2</sub> .....	64
<b>Figura 30-4.</b>	Regresión lineal simple Casos Covid – X <sub>2</sub> _SO <sub>2</sub> .....	65
<b>Figura 31-4.</b>	Regresión lineal simple Casos Covid – X <sub>3</sub> _CO .....	66
<b>Figura 32-4.</b>	Regresión lineal simple Casos Covid – X <sub>2</sub> _SO <sub>2</sub> .....	67

<b>Figura 33-4.</b>	Selección de mejor modelo.....	67
<b>Figura 34-4.</b>	Selección de mejor modelo análisis de Estimados.....	68
<b>Figura 35-4.</b>	Matriz de correlaciones.....	70
<b>Figura 36-4.</b>	Correlación de Pearson CASOS COVID y X1_CO2.....	71
<b>Figura 37-4.</b>	Correlación de Pearson CASOS COVID y X2_O3 .....	71
<b>Figura 38-4.</b>	Correlación de Pearson CASOS COVID y X3_SO2 .....	71
<b>Figura 39-4.</b>	Correlación de Pearson CASOS COVID y X4_CO.....	71
<b>Figura 40-4.</b>	Valores de regresión lineal múltiple .....	72

## ÍNDICE DE ANEXOS

ANEXO A: CÓDIGO MATLAB ANÁLISIS REGRESIÓN LINEAL SIMPLE

ANEXO B: CÓDIGO MATLAB REGRESIÓN LINEAL MÚLTIPLE

ANEXO C: SHAPIRO TEST

ANEXO D: PRUEBA DE NORMALIDAD

ANEXO E: IDEPENDENCIA DE LOS ERRORES

ANEXO F: ANALISIS DE CORRELACION ENTRE DOS VARIABLES

ANEXO G: GRAFICO CENTRAL CON GGLOT CORRELACIONES

ANEXO H: ADICION DE DISTRIBUCIONES

ANEXO I: REGRESION LINEAL MULTIPLE

ANEXO J: CREACION DE MODELO

ANEXO K: SELECCION DE MODELO

## RESUMEN

El presente trabajo de investigación se enfoca en hacer un estudio de la relación de contagios Covid - 19 con la incidencia de contaminantes ambiental PM 2.5, PM10, SO2, O3, CO, CO2; se hace primordial este análisis para tomar decisiones correctas; nuestro análisis dejará proyectar una necesidad de control en un menor tiempo, tomando en cuenta los contaminantes ambientales. El problema que abordamos es tal, que apenas disponemos de datos recientes o anteriores, debemos llevar a cabo un modelo capaz de ofrecer solución al problema propuesto, de tal forma que nuestra actuación en el futuro sea la más acertada. Una vez que hablamos de aprender procesos en los cuales se observa continuidad en el tiempo y en esta situación lo conveniente es hacer una creación de modelos matemáticos sabiendo que es uno de los instrumentos usadas en la actualidad, para el análisis de inconvenientes en la zona de la salud, entre otras; sus fines son explicar, describir y comprobar procesos. Es fundamental nombrar que los modelos matemáticos para patologías infecciosas se aplican como instrumento para apreciar en su justa medida los efectos que va a tener, difícilmente es comprensible un problema complejo sin una mínima modelación, aunque además se debe reconocer que no es viable modelar la integridad de las situaciones reales, esencialmente, la funcionalidad central de producir y examinar los modelos matemáticos es optimizar la comprensión de un procedimiento para prevenir futuras situaciones de patologías, implantar la prevalencia e incidencia y colaborar a tomar elecciones objetivas para el control de enfermedades.

**PALABRAS CLAVE:** <ANÁLISIS DE LA RELACIÓN>, <MEDIO AMBIENTE>, <MODELO MATEMÁTICO>, <INFORMACIÓN>, <ENFERMEDAD VIRAL>, <CONTINUIDAD EN EL TIEMPO >, < ENFERMEDADES INFECCIOSAS >, <PREVALENCIA>, <INCIDENCIA >.



Firmado electrónicamente por:  
**LUIS ALBERTO  
CAMINOS  
VARGAS**



14-11-2022

0169-DBRA-UPT-IPEC-2022



## **SUMMARY / ABSTRACT**

The current research work focuses on a study of the relationship of Covid - 19 infections with the incidence of environmental pollutants PM 2.5, PM10, SO2, O3, CO, CO2. This analysis is essential to make correct decisions. This analysis will allow a need for control to be projected in a shorter time, considering environmental pollutants. The problem we are addressing is such that hardly we have recent or previous data, we must carry out a model capable of offering a solution to the proposed problem, in such a way that our performance in the future could be the most effective. Once we talk about learning processes in which one observes continuity in time and in this situation, it is advisable to create a mathematical model. Knowing that it is one of the instruments currently used for the analysis of problems in health, and other areas; its purpose is to explain, describe and check processes. It is essential to mention that the mathematical models for infectious pathologies are applied as an instrument to appreciate in its fair measure the effects that it will have, it is difficult to understand a complex problem without a minimal modeling, although it must also be recognized that it is not feasible to model the integrity of real situations, essentially, the core functionality of producing and examining mathematical models is to optimize the understanding of a procedure to prevent future situations of pathologies, implement the prevalence and incidence also to collaborate in making objective choices for disease control.

**KEY WORDS:** <RELATIONSHIP ANALYSIS>, <ENVIRONMENT>, <MATH MODEL>, <INFORMATION>, <VIRAL DISEASE>, <CONTINUITY IN TIME >, < INFECTIOUS DISEASES >, <PREVALENCE>, <INCIDENCE>.

# CAPÍTULO I

## 1. INTRODUCCIÓN

### 1.1. Situación problemática

El primer caso se describió el 8 de diciembre de 2019, el 7 de enero de 2020 el Ministerio de Salud de China identificó un nuevo coronavirus que podría ser la etiología, al 2 de enero en China había 835 casos reportados (53 de Hubei). En cuestión de semanas se había extendido a diferentes partes de China. El 13 de enero se reportó el primer caso en Tailandia, el 19 de enero en Corea del Sur, y luego en numerosos países de mundo, debido a lo cual la Organización Mundial de la Salud (OMS), declara desde marzo 2020 como una nueva pandemia mundial. En China todavía no se conoce el origen del virus, aunque se atribuye al pangolín, mamífero usado como alimento. Un brote en curso de un nuevo coronavirus (SARS-CoV-2) ha generado preocupación mundial. Se identifica como causa de neumonía de etiología desconocida. Desde el comienzo del brote en WUHAN, CHINA, posteriormente se ha extendido a todas las provincias de China y a muchos otros países.

La urgente situación epidémica ha estimulado el desarrollo de medicamentos y vacunas antivirales. Como proveedor de servicios líder en el campo de la investigación biológica y el descubrimiento de anticuerpos terapéuticos, detección de candidatos o fármacos. CREATIVE BIOLABS ofrece servicios rápidos y elaborados de descubrimiento de anticuerpos terapéuticos, detección de candidatos o fármacos y desarrollo de vacunas para ayudar a combatir este brote. La clasificación del SARS-CoV-2 usando secuencia imparcial y aislamiento de muestra de pacientes, el SARS-CoV-2 se identificó como otro género BETACORONAVIRUS, familia del CORONAVIRIDAE, los pacientes con SARS-CoV-2 también exhibieron síntomas de neumonía viral que incluyen fiebre, disnea, infiltración pulmonar bilateral, insuficiencia respiratoria e incluso la muerte.

El virus SARS-CoV-2 es extremadamente contagioso, se transmite rápidamente de persona a persona a través de la tos o secreciones respiratorias, por contactos cercanos; las gotas respiratorias de más de cinco micras son capaces de transmitirse a una distancia de hasta dos metros, las manos o los fómites contaminados con secreciones seguido del contacto con la mucosa de la boca, nariz u ojos. Es importante manifestar

que no hubo un aislamiento social a tiempo en China, luego en Italia y España, el virus se esparció rápidamente a diferentes países porque es muy contagiosa.

El nuevo virus afecta directamente al sistema respiratorio, una vez que penetra genera una respuesta inmune anormal de tipo inflamatorio con incremento de citoquinas, esto agrava al paciente, causando daño multiorgánico. Es de la familia de los viejos virus coronavirus, estas dos cepas antiguas causan la gripe común, pero en el 2003 surge la primera mutación, el SARS que inicia en China, con más de 8 460 pacientes en 27 países y una letalidad de 10%, en el año 2012 apareció una nueva cepa mutante de coronavirus en Arabia Saudita, el MERS-CoV, con más de 2499 enfermos y una letalidad del 37%. Ahora se busca una revisión actualizada de aspectos básicos, fundamentales del nuevo coronavirus, el Covid-19, su epidemiología, clínica y terapia

## **1.2. Formulación del problema**

¿Qué modelo matemático permite el estudio de la relación entre la transmisión del Covid-19 y los contaminantes ambientales en la ciudad de Quito-Ecuador?

## **1.3. Situación del problema**

La insuficiencia y desconocimiento para orientar el trabajo y de esta forma obtener resultados seguros, ha implicado examinar las prácticas matemáticas de esta manera verificar el avance de la pandemia, como es el caso de algunos países no desarrollados que no han podido determinar las causas de los contagios, es para esto que debe existir un adecuado análisis de los datos que se tienen hasta el día de hoy, de esta manera obtener un modelo que permita determinar los motivos de contagios.

En la actualidad los organismos de salud de todo el mundo están encargados del tratamiento y en lo particular cuidar de los pacientes críticos contagiados con COVID 19; en Ecuador algunos Hospitales han colapsado, por no saber a qué se está enfrentado, o que tiempo durara esta pandemia, es importante manifestar que el virus está mutando constantemente, es por eso que organismos nacionales e internacionales buscan la solución a esta enfermedad, que nos ha llevado al punto de depender de protección extra para detener el avance de esta pandemia, los organismos de salud que controla y emite los procedimientos, soluciones con el objetivo de detener el avance, al conocer los diferente factores que hacen que la enfermedad alcance su máximo nivel de contagio, al tener un modelo que permita incrementar el fortalecer nuestro sistema de salud pública.

La confusión y problema que produce esta pandemia es de carácter urgente, es decir se busca la solución y muchas veces no se encuentra una salida viable a este problema de salud, ya sea por los escasos recursos que cuenta el país, al no saber cómo controlar o como se origina provoca que los escasos recursos sean destinados de forma aleatoria. Esto hace necesario buscar, investigar y dar una solución, partiendo del inicio y dando seguimiento al caso, considerando las variables necesarias para obtener los resultados requeridos, para que los recursos sean dirigidos a los puntos estratégico, de esta manera dar una gestión y cumpliendo un Plan Operativo que llegue a dar solución a los inconvenientes provocados por el manejo ineficiente de la pandemia.

Al conocer y tener un modelo referencial se debe tener especial cuidado en el manejo de la información para evitar pérdida de eficacia, de esta forma las entidades de salud, busquen las mejores opciones para controlar y dar soluciones rápidas, es importante tener en cuenta que, se ha detectado el virus muta constantemente haciéndolo más viral, por eso nuestro modelo debe ajustarse a la necesidad actual y a futuro, de ser necesario realizar ciertos cambios por las diferentes variables a considerar.

#### **1.4. Formulación del problema**

¿De qué manera un modelo matemático nos permitirá el estudio de la relación entre la transmisión del Covid-19 y los contaminantes ambientales en la ciudad de Quito-Ecuador?

#### **1.5. Preguntas directrices**

- ¿De qué manera influye la transmisión del COVID-19 y los contaminantes ambientales en la ciudad de Quito-Ecuador?
- ¿Cómo mejorar la eficiencia y eficacia del manejo de la pandemia COVID-19 con un modelo matemático?
- ¿Cómo los modelos matemáticos permiten identificar las variables necesarias para el estudio?
- ¿Cómo podemos modelar situaciones experimentales donde la variable es cambiante y su respuesta es tipo Binomial?
- ¿Cómo podemos validar el modelo planteado?

## 1.6. Justificación

El presente trabajo investigativo se desarrollará en la ciudad de Quito, con los datos manejados por el ministerio de SALUD PUBLICA, GESTIÓN DE RIESGOS Y LA RED DE MONITOREO ATMOSFÉRICO DEL DISTRITO METROPOLITANO DE QUITO

Este proyecto de investigación se centra en la ciudad de QUITO, la aplicación de un modelo matemático que permite relacionar los contagios y los contaminantes en la ciudad, en función de obtener resultados óptimos de la misma manera reflejar los logros requeridos en este estudio.

Es considerable manifestar que en las iniciaciones de la pandemia fue perceptible la inexperiencia sobre el impacto que tendría el SARS-CoV-2 en todas las ciudades, está crisis ha generado graves repercusiones a nivel social, económico, salud. En un inicio los primeros contagios fueron en estricta reserva y las complicaciones de los pacientes fueron reservadas, lo que generó, una respuesta tardía por parte de la comunidad científica y el único asidero era el bagaje de conocimientos, esto ha llevado que se realicen múltiples estudios sobre el comportamiento de los coronavirus. Si bien es cierto, es que se aprendió y se sigue aprendiendo sobre la marcha, pero es importante manejar las variables que arrojan el diario vivir para mejorar sustancialmente el índice de contagios en vista que en la actualidad con la vacuna se ha controlado de manera sustancial la salud de la comunidad

Por supuesto, al aplicar modelo matemático para verificar la relación del virus y los contaminantes, sería verificar la evolución del virus y el impacto que tendría, al no ser controlado, todo esto es una intención más o menos académica y ejecutiva al conocer la dinámica nos llevara a unos criterios de validez y verificabilidad sobre todo en el área de la salud, que ameritan tiempo y en ocasiones evitan por criterio ético la inmediatez de la información que es demandada.

No obstante que para mejorar el modelo nos debemos concentrar en los parámetros fundamentales, como el número de contagios y la relación con los contaminantes, con estas dos variables nos permitirá medir la relación. Los modelos parten de un valor para cuando el virus circular libremente, en condiciones normales, es decir parámetros de ambiente estándar antes de haber tomado medidas. Esto depende de

factores los cuales harán que el índice de contagios aumente considerablemente y, por tanto, será muy distinto en zonas de baja contaminación, o donde las personas viven aisladas y apenas se tocan, con en zonas agrícolas o el campo, esto se ralentiza en las grandes urbes donde.

Esta investigación es pertinente y relevante debido que es oportuna y conveniente es decir, el tema propuesto tiene un argumento real el cual va a aportar a la comunidad de manera positiva y se puede argumentar la razón que la investigadora pretende demostrar que se incorporan datos reales los mismos que nos arrojaran resultados viables esto puede ayudar en las fases de estimación o ajuste antes de la intervención, pero no resuelve la fase de predicción porque necesita asumir hipótesis sobre la reducción de índices de contaminación, conjuntamente a las condiciones en las que el virus muta. La parte teórica se realiza de manera crítica y reflexiva, de tal manera, que sea referente para futuras investigaciones de este tipo.

Es importante manifestar que existirán otros modelos de gran interés los mismo que permitirán enfrentar este tipo de crisis, la misma que se ha vuelto mundial dinámicamente al trabajar con una investigación que para muchos es desconocida, nos dejara una enseñanza, dejando de una manera efectividad una base firme contra la expansión del virus. Asimismo, este trabajo nos dará la iniciativa “**Acción matemática contra el coronavirus**” contribuye al conocimiento de la epidemia desde las matemáticas.

## **1.7. Objetivos**

### ***1.7.1. Objetivo general***

Desarrollar un modelo matemático para relacionar el número de contagios del covid-19 y los contaminantes ambientales en la ciudad de Quito Ecuador. Periodo enero 2020 – marzo 2021.

### ***1.7.2. Objetivos específicos***

- a) Recopilar los datos de los casos de infectados diarios por Covid-19 en la ciudad de Quito-Ecuador en el periodo comprendido del 13 de marzo 2020 al 13 de marzo 2021.

- b) Construir la base de datos de los contaminantes del aire PM2.5, PM10, O3, SO2 y CO para la ciudad de Quito-Ecuador en el periodo comprendido del 13 de marzo 2020 al 13 de marzo 2021.
- c) Establecer el modelo matemático entre el número de casos nuevos diarios por Covid-19 y cada contaminante PM2.5, PM10, O3, SO2 y CO.

## **1.8. Hipótesis**

### ***1.8.1. Hipótesis general***

H0: Las variables independientes (PM2.5, PM10, O3, SO2 y CO) realizan un aporte significativo al modelo matemático, en donde se relaciona el número de contagios COVID -19 y los contaminantes ambientales

H1: Las variables independientes (PM2.5, PM10, O3, SO2 y CO) no realizan un aporte significativo al modelo matemático, en donde se relaciona el número de contagios COVID -19 y los contaminantes ambientales

## **1.9. Identificación de variables**

- **Variable independiente:** Contaminantes ambientales (PM2.5, PM10, O3, SO2 y CO).
- **Variable dependiente:** Transmisión del Covid-19.

## CAPÍTULO II

### 2. MARCO TEÓRICO

#### 2.1. Antecedentes del problema

Al revisar varios trabajos de investigación que relacionan el trabajo de tesis y nutren a este trabajo, se tiene que:

El 30 de diciembre del 2019 se identifican los primeros casos de neumonía por causa desconocida y se la reportada en la Comisión Nacional de Salud de China, los pacientes presentaban como antecedente epidemiológico haber asistido al mercado de mariscos y animales de Wuhan, ya sea siendo trabajadores o visitantes. El 1 de enero de 2020 se decide cerrar el mercado y el 30 de enero la Organización Mundial de la Salud (OMS) declara emergencia de salud pública de interés internacional (ESPII) (34). El día 7 de enero, se identifica por primera vez el coronavirus. Posteriormente, el virus se denominó oficialmente síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2), la enfermedad que lo causa se conoce como COVID-19. El 11 de enero se reportó el primer deceso. Publicaciones recientes indican que el SARS-CoV-2 pertenece al género Betacoronavirus, con más del 85% de identidad.

Este virus comparte un ancestro común con los virus mortales del Síndrome Respiratorio Agudo Severo (SARS) y el Síndrome Respiratorio del Medio Oriente (MERS). Según el Center for Systems Science, el número de personas infectadas con el virus ha aumentado exponencialmente hasta la fecha (31,48), con 60.364 casos confirmados de COVID-19 y 1.370 muertes reportadas al 13 de febrero de 2020 Razón Johns Hopkins Ingeniería e Ingeniería 2. Los primeros casos reportados fuera de China ocurrieron en Tailandia el 13 de enero, Japón el 16 de enero y Corea del Sur el 19 de enero. La sintomatología más frecuente presentada en las personas con COVID-19 es fiebre, tos, fatiga y dificultad respiratoria.

- Fiebre: entre el 81-98% los pacientes presentaron fiebre, se ha informado que los pacientes menos graves suelen tener fiebre baja, mientras que los casos más graves presentan fiebre de baja a moderada, o incluso no tenerla.



- Tos: entre el 48,2%- 59% los pacientes presentan tos, a menudo se manifiesta como tos seca irritante con o sin esputo, sin embargo, SARS-CoV-2 invade las células epiteliales alveolares, por lo que, los síntomas como la tos no son particularmente evidentes o puede no presentarse.
- Dificultad para respirar: la disnea se ha reportado en un 31% de los pacientes. Algunos pacientes se quejaron de opresión en el pecho o disnea debido a la tensión al momento de la consulta.

Fatiga: entre el 32,1%- 69% de los pacientes se sintieron cansados y débiles, y en algunos casos acompañado de mialgia (35%). Adicionalmente, existen algunos pacientes con sintomatologías inicial atípica como: digestiva (anorexia leve, náuseas, vómitos, diarrea, dolor abdominal, etc.); neurológicas (dolor de cabeza); cardiovascular (como palpitaciones, opresión en el pecho, etc.); o sólo dolor leve en las extremidades o mialgia dorsal, además del reporte de síntomas oftalmológicos como primera manifestación (conjuntivitis) (40,63). Finalmente, hay reportes de infecciones asintomáticas. Respecto a niños y niñas, se ha descrito un cuadro similar, sin embargo, puede haber variaciones en función del peso de estos, también se pueden dar casos asintomáticos o de progresión rápida (5,29). El reservorio animal del SARS-CoV-2 aún se desconoce, no se había encontrado antes en humanos y el conocimiento de cómo se propaga todavía se encuentra en evolución. Sin embargo, algunos de los reportes han mencionado que las primeras personas infectadas trabajaron o visitaron el mercado de mariscos de Wuhan en China. Además de los mariscos, se informa que en este lugar se vendieron serpientes, pájaros y otros pequeños mamíferos, incluidas marmotas y murciélagos. Hasta la fecha el reservorio más probable es el murciélago (23,27). La OMS informó que las muestras ambientales tomadas del mercado han dado positivo para el nuevo coronavirus, pero no se ha identificado una asociación animal específica.

A nivel mundial los investigadores están trabajando actualmente para identificar el reservorio de SARS-CoV-2, los posibles reservorios animales intermedios (3,49). Ninguno de los casos exportados tuvo contacto con dicho mercado, lo que sugiere una transmisión de persona a persona. La vía de transmisión es a través de las vías respiratorias y por contacto con personas infectadas o fómites con el virus. La vía de transmisión fecal-oral aún no se ha determinado, sin embargo, algunos equipos han implementado medidas de prevención por esta vía. Se ha reportado como periodo de

incubación (periodo entre la infección y la manifestación clínica o aparición de síntomas) ha variado a medida que progresa la investigación (12,33,73), se ha establecido un periodo de entre 1 a 14 días. Los casos graves progresan de 7 a 10 días después del inicio de COVID-19, lo que sugiere que la inmunidad adaptativa está involucrada en el daño inmunológico durante la progresión de la enfermedad. La transmisión del virus generalmente se evalúa mediante el número de reproducción del caso cero ( $R_0$ ). La OMS considera que la capacidad de transmisión de SARS es fuerte y un artículo que revisa los cálculos de  $R_0$  disponibles a la fecha (incluido el de OMS) señala que tiene una moderada transmisibilidad. Un reciente cálculo del ritmo reproductivo básico del virus elaborado por el Centro de Control y Prevención de Enfermedades de China, CDC por sus siglas en inglés (Center for Disease Control and Prevention) lo estimó en 2,2 (IC 95%: 1,4 a 3,9) (8,59), sin embargo, las estimaciones varían dependiendo del modo de cálculo, en contexto y a medida que se implementa el manejo efectivo de la infección. De acuerdo con los motivos que se pueden explicar la alta capacidad de propagación del SARS-CoV-2 es que sería transmisible durante el período de incubación, que existen casos con sintomatología atípica o asintomáticos que dificultan su diagnóstico, y que el brote coincidió con festividades en China.

Con respecto a su forma de actuar, el equipo de la Universidad de Tongji en Shanghai cree que SARS-CoV-2 parece haberse convertido en un virus que puede secuestrar las células epiteliales alveolares tipo II para facilitar la proliferación y transmisión del virus. En rigor toda la población es susceptible a enfermar de COVID-19, sin embargo, varias publicaciones destacan que personas de edad avanzada y/o con comorbilidades crónicas pueden ser más susceptible a infectarse o a presentar manifestaciones de la enfermedad más graves (3,46,49,69,92), estos pacientes presenten ambas características o solo una. En el grupo de los primeros 425 casos confirmados, la edad promedio fue de 59 años, y el 56% eran hombres.

De acuerdo con las comorbilidades se han indicado hipertensión, diabetes, obesidad, enfermedades cardiovasculares o renales. Se está considerando que SARS-CoV-2 parece tener un potencial patogénico similar al SARS-CoV y MERS-CoV, una infección por SARS-CoV-2 en periodo de gestación podría estar asociado a manifestaciones más graves de la enfermedad, tales como muerte materna o neonatal y complicaciones en el recién nacido.

En el estudio, los resultados indican que la vigilancia es necesaria para estos pacientes vulnerables después de la infección con SARS-CoV-2. Respecto a si hay mayor susceptibilidad en hombres, este patrón se observó en algunos brotes, sin embargo, una de las hipótesis es que en el lugar de exposición del primer brote había más personas de este sexo. En los casos de revisión pediátrica, una publicación señala que, según la experiencia con otros coronavirus y la situación epidemiológica actual de China, se espera que la susceptibilidad de niñas y niños al SARS-CoV-2 sea menor que la de personas adultas.

El uso de modelos matemáticos para predecir cómo se comportará la enfermedad infecciosa COVID-19, el nivel de contagio que puede alcanzar y cuánto durará, esto nos permitirá jugar un papel importante en el control de infecciones. El efecto es saludable y normal. vida cotidiana de los ciudadanos. Se presenta el primer modelo desarrollado para el COVID-19. Los supuestos generales considerados para el desarrollo del modelo son los siguientes:

Utilizar modelos matemáticos para predecir cómo se desarrollará el comportamiento de un contagio de COVID-19, el nivel de contagio que puede alcanzar y cuánto tiempo se desarrollará y jugará un papel fundamental en el control de contagios, reabriendo, utilizando los mejores recursos y servicios del sector salud con normalidad. vida cotidiana de los ciudadanos. Introdujo los primeros modelos desarrollados para COVID-19. Los supuestos generales considerados para el desarrollo del modelo son los siguientes:

- El modelo fue construido a partir de las series de tiempo de casos confirmados de infección por covid-19 reportadas por el Ministerio de Salud Pública, gestión de riesgos y la red de monitoreo de atmosférico distrito metropolitano de Quito.
- Los casos confirmados en la ciudad de Quito se refieren a los pacientes que se contagiaron por causas ambientales, mientras que los casos relacionados corresponden a pacientes que tuvieron contacto con las personas infectadas en el país.
- Los casos confirmados de infección por COVID-19 reportados como asilados presentan un comportamiento sigmoidal.

- Los casos confirmados de infección por COVID-19 reportados como casos relacionados presentan un crecimiento exponencial inicial asociado a un brote no controlado de la infección.
  - Los casos confirmados relacionados de infección por COVID-19 son proyectados, adicionalmente, a través de una función algebraica polinomial cúbica. Bases Teóricas

### **Estadística Descriptiva**

La estadística descriptiva es la rama de las Matemáticas que recolecta, presenta y caracteriza un conjunto de datos con el fin de describir apropiadamente diversas características del conjunto.

- Al conjunto de los distintos valores numéricos que adopta un carácter cuantitativo se llama variable estadística las variables pueden ser de dos tipos: Variables cualitativas o categóricas: no se pueden medir numéricamente (por ejemplo: nacionalidad, color de la piel, sexo).
- Variables cuantitativas: tienen valor numérico (edad, precio de un producto, ingresos anuales).

La estadística para las variables cuantitativas es:

- Media
- Mediana
- Moda
- Varianza
- Desviación estándar

La estadística para las variables cualitativas es:

- Tabla de frecuencias
- Tablas de contingencia

### 2.1.1 Modelos de regresión línea

#### 2.1.1.1 Regresión lineal simple

La regresión lineal consiste en generar una ecuación (modelo) que, basándose en la relación existente entre ambas variables, permita predecir el valor de una a partir de la otra. El modelo varía según qué variable se considere dependiente de la otra (lo cual no implica causa-efecto). (Amat Rodrigo, 2016).

Según Zurita Herrera (2010), la regresión lineal simple consiste en generar un modelo de regresión (ecuación lineal) que pueda explicar la relación lineal que existe entre dos variables. La variable dependiente o de respuesta se identifica como Y, y la variable predictora o independiente se identifica como X. Un modelo de regresión lineal simple se describe de acuerdo con la siguiente ecuación:

$$Y_i = Q_0 + Q_1X_i + s_i \quad (1)$$

$\beta_0 =$  Es la ordenada en el origen

$\beta_1 =$  Pendiente

$\varepsilon_i =$  Error aleatorio

$X_i =$  Variable de explicación, condiciona el valor de Y

Dado el modelo condicional se trabaja en los siguientes supuestos:

$$E(Y_i|X = X_i) = B_0 + B_1X_i \quad (2)$$

$$E(s_i) = 0 \quad \text{Normalidad} \quad (3)$$

$$Var(s_j) = \sigma^2 \quad \text{Varianza Constante} \quad (4)$$

$$cov(s_i, s_j) = 0; \quad i \neq j \quad \text{Independencia} \quad (5)$$

$$cov(s_i, s_j) = 0; \quad i \neq j \quad \text{Independencia} \quad (6)$$

Dónde:

$E(Y_i|X = X_i) =$  Parte sistemática o determinística del modelo

Los valores  $B_0$ ,  $B_1$  y  $\sigma^2$ , son constantes desconocidas, pero estadísticamente estimables.

El hecho de que la varianza  $\sigma^2$  del error sea constante durante todo el proceso es un supuesto fuerte y hace que el modelo utilizado sea considerado como homocedástico, es decir que la variabilidad es constante (Zurita Herrera, 2010). El error aleatorio representa la diferencia entre el valor ajustado por la recta y el valor real. Recoge el efecto de todas aquellas variables que influyen en  $Y$  pero que no se incluyen en el modelo como predictores (Amat Rodrigo, 2016). En la gran mayoría de casos, los valores  $\beta_0$  y  $\beta_1$  poblacionales son desconocidos, por lo que, a partir de una muestra, se obtienen sus estimaciones como:

$$\hat{y} = \hat{Q} + \hat{Q}x \quad (7)$$

Estas estimaciones se conocen como coeficientes de regresión o least square coefficient estimates, ya que toman aquellos valores que minimizan la suma de cuadrados residuales, dando lugar a la recta que pasa más cerca de todos los puntos (Amat Rodrigo, 2016).

$$\hat{y} = \hat{Q} + \hat{Q}x \quad (8)$$

$$\hat{Q} = \frac{S_{xy}}{S_{xx}} \quad (9)$$

Dónde:

$$S_{xx} = \frac{\sum(x_i - \bar{x})^2}{n - 1} ; \quad S_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

Dando como resultado:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Mientras que  $\hat{\beta}_0$  es:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Dónde  $S_{xx}$  y  $S_{xy}$  son las desviaciones típicas de cada variable.  $\hat{Q}$  es el valor esperado la variable  $Y$  cuando  $X = 0$ , es decir, la intersección de la recta con el eje  $y$ . Una

recta de regresión puede emplearse para diferentes propósitos y dependiendo de ellos es necesario satisfacer distintas condiciones. Siendo  $\bar{x}$  y  $\bar{y}$  las medias aritméticas de los valores observados de X y de Y;  $\hat{y}$  es el valor que el modelo estima para  $Y_i$ , dado  $X = x_i$  con lo efectuado anteriormente se puede obtener el coeficiente de Correlación  $r_{xy}$  siendo:

$$s_{yy} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} \quad (10)$$

Dando como resultado:

$$r_{xy} = \frac{\hat{Q} \sqrt{s_{xx}}}{s_{yy}} \quad (11)$$

Por otro lado, el Coeficiente de determinación  $R^2$  se define como el cociente de la suma cuadrática de Regresión para la suma Total:

Donde la suma cuadrática de Regresión (SCR) se define de la siguiente manera:

$$SCR = \sum_{i=1}^n (\hat{y} - \bar{y})^2 \quad (12)$$

La suma cuadrática del Error (SCE) se define de la siguiente manera:

$$SCE = \sum_{i=1}^n (y_i - \hat{y})^2 \quad (13)$$

La suma cuadrática Total (SCT) se define de la siguiente manera:

$$SCT = SCR + SCE \quad (14)$$

$$SCT = \sum_{i=1}^n (\hat{y} - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y})^2 \quad (15)$$

Dando como resultado:

$$R^2 = \frac{SCR}{SCT} = \frac{SCT - SCE}{SCT} = 1 - \frac{SCE}{SCT} \quad (16)$$

Por otro lado, se puede comprobar en la regresión lineal simple la relación entre el coeficiente de Correlación de X con Y y el coeficiente de determinación que viene dado por:

$$r_{xy} = \pm\sqrt{R^2} \quad (17)$$

De acuerdo Zurita Herrera (2010) menciona el teorema de Cochran que establece que si cada una de las n observaciones  $y_i$  son tomadas de una misma Población Normal con parámetros  $\mu_i$  y  $\sigma^2$ , la SCT se descompone en K sumas Cuadráticas que se denota  $SC_q$ ,  $q = 1; 2; \dots; k$ . Cada una de ellas con  $gl_q$  grados de libertad; por lo que, el cociente  $SC_q/\sigma^2$  tiene distribución Ji-Cuadrado con  $gl_q$  grados de libertad, teniendo la SCT, (n-1) grados de libertad esto significa:

$$\sum_{q=1}^k gl_q = n - 1 \quad (18)$$

Se ha efectuado una partición de la SCT, donde  $K = 2$ , la una parte de la SCT es la SCR ( $q=1$ ) y la otra es la SCE, con  $gl_1 = (p - 1)$  y la SCE ( $q = 2 = k$ ) con  $gl_2 = (n - p)$ .

En la tabla 1-2 se visualiza la tabla de análisis de varianza o ANOVA que consiste en un arreglo rectangular cuyos componentes son: las fuentes de variación, grados de libertad, las sumas y medias cuadráticas y un valor adicional que es el estadístico de Prueba F.

Tabla 1-2 Análisis de Varianza

Fuentes de Variación	Grados de libertad	Sumas Cuadráticas	Medias Cuadráticas	Estadístico de Prueba F
Regresión	$p - 1$	$\sum_{i=1}^n (y_i - \bar{y})^2$	$\frac{SCR}{p - 1}$	$F_0 = \frac{MCR}{MCE}$
Error (residuos)	$n - p$	$\sum_{i=1}^n (y_i - \hat{y})^2$	$\frac{SCE}{n - p}$	



---

Total	$n - 1$	$\sum_{i=1}^n (y_i - \bar{y})^2$
-------	---------	----------------------------------

---

Fuente: Adaptado de Zurita Herrera, (2010)

### 2.1.1.2 Distribución F de Fisher

Es el cociente de:

$$F_0 = \frac{MCR}{MCE} = \frac{SCR/1}{SCE/n(n-2)} \quad (19)$$

Es una variable aleatoria F con  $(p - 1) = 1$  gl en el numerador y  $(n - p)$  gl en el denominador. Si la Hipótesis Nula es verdadera ( $H_0: B_1 = 0$ ) entonces,  $E(MCR) = E(MCE) = \sigma^2$ , por lo que, un valor del estadístico de prueba F, cercano a “uno” sería evidencia de que de que la Hipótesis Nula es verdadera, en tanto que un valor alejado de uno evidencia la necesidad de rechazar  $H_0$ .

Por lo tanto, establecido el contraste:

$$H_0: Q_1 = 0 \text{ Vs. } H_1: Q_1 \neq 0 \quad (20)$$

Con  $(1 - \alpha)$  100% de confianza la Hipótesis Nula debe ser rechazada si el Estadístico de Prueba:

$$F_0 = \frac{MCR}{MCE} > F_{(\alpha; p-1, n-p)} \quad (21)$$

Una recta de regresión puede emplearse para diferentes propósitos y dependiendo de ellos es necesario satisfacer distintas condiciones. Sin embargo, en caso de querer predecir el valor de una variable en función de la otra, no solo se necesita calcular la recta, sino que además hay que asegurar que el modelo se ajusta a los datos (Amat Rodrigo, 2016).

### 2.1.1.3 Inferencia de Regresión lineal para $\beta_0$ y $\beta_1$

Significa que el modelo generado es una estimación de la relación poblacional a partir de la relación que se observa en la muestra y, por lo tanto, está sujeta a variaciones. Para cada uno de los parámetros de la ecuación de regresión lineal simple ( $\beta_0$  y  $\beta_1$ ) se

puede calcular su significancia (p-value) y su intervalo de confianza. La prueba estadística más empleada es el t-test (existen alternativas no paramétricas).

La prueba de significancia para la pendiente ( $\beta_1$ ) del modelo lineal considera como hipótesis:

H0: No hay relación lineal entre ambas variables, por lo que, la pendiente del modelo lineal es cero.  $\beta_1 = 0$

H1: Sí hay relación lineal entre ambas variables por lo que la pendiente del modelo lineal es distinta de cero.  $\beta_1 \neq 0$

#### 2.1.1.4 Los intervalos de confianza para cada $Q_1$

Con  $(1-\alpha)$  100% de confianza para  $\beta_i$ , el intervalo para cada  $\beta_i$  es:

$$\hat{Q} \pm SE(\hat{Q}) t_{(\frac{\alpha}{2}, n-2)} \quad (22)$$

El intervalo quedaría como:

$$\hat{Q} - SE(\hat{Q}) t_{(\frac{\alpha}{2}, n-2)} \leq Q_i \leq \hat{Q} + SE(\hat{Q}) t_{(\frac{\alpha}{2}, n-2)} \quad (23)$$

donde  $t_{(\frac{\alpha}{2}, n-2)}$  es el percentil de la distribución t de Student con n-2 grados de libertad que deja a su derecha un área de  $\alpha/2$ . Cuanto menor es el número de observaciones n, menor la capacidad para calcular el error estándar del modelo. Como consecuencia, la exactitud de los coeficientes de regresión estimados se reduce.

#### 2.1.1.5 Regresión lineal múltiple

Permite generar un modelo lineal en el que el valor de la variable dependiente o respuesta (Y) se determina a partir de un conjunto de variables independientes llamadas predictores ( $X_1, X_2, X_3 \dots$ ). Es una extensión de la regresión lineal simple, por lo que es fundamental comprender esta última. Los modelos de regresión múltiple pueden emplearse para predecir el valor de la variable dependiente o para evaluar la influencia que tienen los predictores sobre ella (esto último se debe que analizar con cautela para no malinterpretar causa-efecto) (Amat Rodrigo, 2016).

Los modelos lineales múltiples siguen la siguiente ecuación:

$$Y_i = Q_0 + Q_1 X_{1i} + Q_2 X_{2i} \dots + Q_n X_{ni} + S_i \quad (24)$$

dónde

$$\varepsilon_i \sim N(0, \sigma^2 I_n) \quad \text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j; i, j = 1, 2, \dots, n$$

$\beta_0$ : es la ordenada en el origen, el valor de la variable dependiente Y cuando todos los predictores son cero.

$\beta_i$ : es el efecto promedio que tiene el incremento en una unidad de la variable  $X_i$  sobre la variable dependiente Y, manteniéndose constantes el resto de las variables. Se conocen como coeficientes parciales de regresión.

$\varepsilon_i$  : es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo.

Esta vez con los parámetros  $\beta_0, \beta_1, \beta_2$ , dos variables de explicación  $X_1$  y  $X_2$  y el error  $\varepsilon_i$ . Es importante tener en cuenta que la magnitud de cada coeficiente parcial de regresión depende de las unidades en las que se mida la variable predictora a la que corresponde, por lo que su magnitud no está asociada con la importancia de cada predictor. Para poder determinar qué impacto tienen en el modelo cada una de las variables, se emplean los coeficientes parciales estandarizados, que se obtienen al estandarizar (sustraer la media y dividir entre la desviación estándar) las variables predictoras previo ajuste del modelo (Amat Rodrigo, 2016).

### 2.1.1. Ajuste por mínimos cuadrados

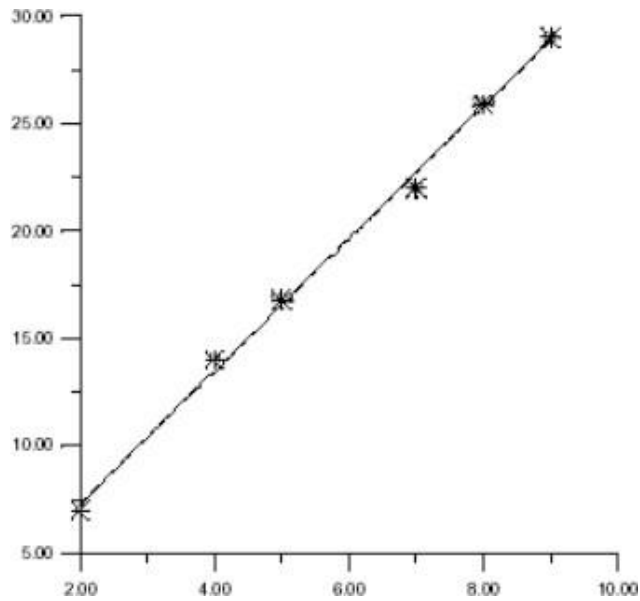
Existen numerosas leyes físicas en las que se sabe de antemano que dos magnitudes  $x$  e  $y$  se relacionan a través de una ecuación lineal.

$$y = ax + b \tag{25}$$

donde las constantes  $b$  (ordenada en el origen) y  $a$  (pendiente) dependen del tipo de sistema que se estudia y, a menudo, son los parámetros que se pretende encontrar. Por ejemplo, la fuerza  $F$  de tracción sobre un muelle y el alargamiento  $I$  que experimenta éste está ligadas a través de una ley lineal:

$$I = (1/K)F \tag{26}$$

con ordenada en el origen cero y donde el inverso de la pendiente ( $K$ ) es una característica propia de cada muelle: la llamada constante elástica del mismo.



**Figura 1-2.** Métodos mínimos cuadrados

Fuente: (Amat Rodrigo, 2016).

El método más efectivo para determinar los parámetros  $a$  y  $b$  se conoce como técnica de mínimos cuadrados. Esta consiste en someter el sistema a diferentes condiciones, fijando para ello distintos valores de la variable independiente  $x$ , y anotando en cada caso el correspondiente valor medido para la variable dependiente  $y$ . De este modo se dispone de una serie de puntos  $(x_1, y_1)$ , ...,  $(x_n, y_n)$  que, representados gráficamente, deberían caer sobre una línea recta. Sin embargo, los errores experimentales siempre presentes hacen que no se hallen perfectamente alineados (ver Fig. 2-3). El método de mínimos cuadrados determina los valores de los parámetros  $a$  y  $b$  de la recta que mejor se ajusta a los datos experimentales. Sin detallar el procedimiento, se dará aquí simplemente el resultado:

$$a = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2} \quad (27)$$

$$b = \frac{(\sum y_i) - a(\sum x_i)}{n} \quad (28)$$

dónde  $n$  es el número de medidas y  $\Sigma$  representa la suma de todos los datos que se indican.

Los errores en las medidas se traducirán en errores en los resultados de  $a$  y  $b$ . Se describe a continuación un método para calcular estos errores. En principio, el método de

mínimos cuadrados asume que, al fijar las condiciones experimentales, los valores  $y_i$  de la variable independiente se conocen con precisión absoluta (esto generalmente no es así, pero lo aceptamos como esencial en el método). Sin embargo, las mediciones de la variable  $x$ , irán afectadas de sus errores correspondientes, si  $\epsilon$  es el valor máximo de todos estos errores, entonces se tiene:

$$\Delta a = \frac{\sqrt{n} s}{\sqrt{n \sum_1^n x_1^2 - (\sum_1^n x_1)^2}} \quad (29)$$

$$\Delta b = \frac{s}{\sqrt{n}}$$

La pendiente de la recta se escribirá  $a \pm \Delta a$ , y la ordenada en el origen  $b \pm \Delta b$ . El coeficiente de correlación es otro parámetro para el estudio de una distribución bidimensional, que nos indica el grado de dependencia entre las variables  $x$  e  $y$ . El coeficiente de correlación  $r$  es un número que se obtiene mediante la fórmula:

$$r = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{[n(\sum x_i^2) - (\sum x_i)^2][n(\sum y_i^2) - (\sum y_i)^2]}} \quad (30)$$

Su valor puede variar entre 1 y -1.

Si  $r = -1$  todos los puntos se encuentran sobre la recta existiendo una correlación que es perfecta e inversa.

Si  $r = 0$  no existe ninguna relación entre las variables.

Si  $r = 1$  todos los puntos se encuentran sobre la recta existiendo una correlación que es perfecta y directa.

Ejemplo:

Supongamos un muelle sometido a tracción, se ha cargado el muelle con diferentes pesos ( $F$ , variable independiente o  $y$ ) y se han anotado los alargamientos ( $l$  variable dependiente o  $x$ ).

Tabla 2-2 Ejemplo de mínimos cuadrados

Cargas sucesivas F(Y <sub>i</sub> ) gramos	Lecturas sucesivas (X <sub>i</sub> ) L mm	
	200	60
	400	120
500	150	
700	210	
900	260	
1000	290	

Los distintos datos que se necesitan son:

Tabla 3-2 Ejemplo de mínimos cuadrados

n	6
$\sum x_i$	1090
$\sum x_i^2$	236300
$\sum y_i$	3700
$\sum y_i^2$	2750000
$\sum x_i y_i$	806000
s	0,2

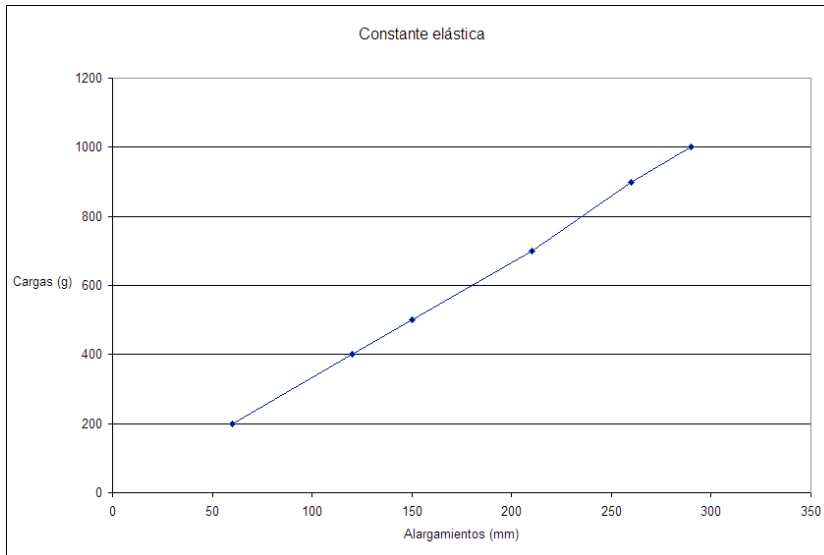
con lo cual aplicando las expresiones [27], [28], [29] y [30]

$$b = -18,4153; a = 3,4959; \Delta b = 0,08164966; \Delta a = 0,00102217; r = 0,9995$$

Redondeando en la forma usual  $b = -18,42 \pm 0,08 \text{ mm}$ ;  $a = 3,50 \pm 0,00 \text{ mm/Kp}$

No se debe olvidar que se persigue el valor de la constante elástica del muelle:

$$K = a = \frac{F}{l} \quad (31)$$



**Figura 2-2.** Ejemplo de mínimos cuadrados

**Fuente:** Adaptación propia del gráfico presentado en Sargent, 2010.

## 2.2. Desarrollo de modelos matemáticos

Es importante manifestar que, para verificar el análisis de un sistema, es necesario obtener un modelo matemático que represente el escenario planteado, con el cual el proceso y desarrollar estrategias de control para luego ser aplicadas en el sistema real. Con este objetivo, es importante que el modelo, no siendo único, sea una representación. En este trabajo se logra un modelo matemático, la elaboración del modelo consta de dos partes, la primera es encontrar las ecuaciones físicas del sistema en términos de algunos parámetros desconocidos, la segunda es hallar estos parámetros usando mínimos cuadrados. Se realizó una validación del modelo obtenido sobre la planta y se encontró que el modelo era confiable.

## 2.3. Supuestos del Modelo

Al realizar un estudio de campo y manejando datos, los mismos que a lo largo del tiempo cambian, cabe mencionar que se indica que un modelo matemático necesita de los siguiente:

### 2.3.1. Validación de Modelos Desde la Perspectiva Epistemológica

La confirmación de modelos en el periodo de vida del modelado y la simulación (M&S) es un asunto de gran importancia para avalar no sólo la calidad de resultados sino

también de la simulación, la eficacia epistémica del modelo obtenido y de los resultados de la experimentación numérica asociados a dicho modelo, como representación correcta, desde la perspectiva de un problema, fundamental o aplicado, se formula a partir de peculiaridades y comportamientos propios de un sistema u objeto del mundo real. Así pues, se debe considerar que varios trabajos abordan el problema del conocimiento que puede ser obtenido mediante el análisis conceptual del modelo para simulación y de los patrones de comportamiento exhibidos por la simulación de ambientes. A continuación, se presenta el estado de la investigación sobre los aspectos epistemológicos referidos a la validación conceptual y aplicada de modelos obtenidos mediante estudios basados en M&S.

El perfeccionamiento de modelos para la representación de un aspecto en particular de la realidad ha estado presente desde las primeras etapas del pensamiento científico. De particular interés es el caso del modelo geocéntrico del universo de Ptolomeo, que abordó el estudio de una realidad física de más difícil acceso a la experiencia directa, y que buscaba, de manera explícita, la coincidencia de las predicciones del modelo con las órbitas, aparentemente caprichosas, de los planetas observables para la época. Estos objetivos -describir, explicar, predecir y facilitar la medida de fenómenos naturales-, sumados al concepto de prueba, introducida por Pitágoras como sustento argumentativo de sus modelos matemáticos (principalmente abstractos), contienen los cimientos de la disciplina del modelado.

De otro lado, el uso del modelado matemático para la solución de problemas del mundo real, específicamente aquellos que tenían mayor relevancia o prioridad por su utilidad práctica, implicaba el desarrollo de métodos reproducibles y de uso accesible por los usuarios finales, de modo que se obtuvieran resultados válidos y útiles.

Tal necesidad fue abordada por Al-Jwarizmi en sus trabajos sobre métodos algebraicos que conducirían posteriormente al desarrollo del concepto de algoritmo, en el contexto de las ciencias de la computación. Los algoritmos son la base sobre la que se construyen implementaciones numéricas de modelos adecuados para la reproducción de características y comportamientos de los fenómenos modelados, mediante el uso del computador, lo que se conoce actualmente como *simulación*. Bunge (1969) definió simulación como una relación entre los objetos  $x$  y  $y$ , donde  $x$  “simula” a  $y$ , si: a) existe una relación de correspondencia entre las partes y las propiedades de  $x$  y  $y$ ; b) la analogía es de valor para  $x$ , o la entidad  $z$  que la controla. Guala (2002) critica esta definición por ser “antropocéntrica” pero, sobre todo, imprecisa, al permitir que incluya, entre otros, la



similitud entre fenómenos no controlados ni observados y el simple uso de analogías en descripciones verbales o la construcción de modelos estáticos

### **2.3.2. *El Modelo***

En su sentido más amplio, un modelo situacional es un conjunto de representaciones formales que contienen conocimiento sobre un problema de interés para la investigación adquirido a través de todas las fuentes relevantes.

De esta forma, el modelo existe en la especificación formal de los elementos del sistema, las relaciones entre ellos y los parámetros que permiten contextualizar la práctica del sistema según las especificidades del entorno y la relación de la realidad sistema del mundo al sistema del mundo real. Así, el desarrollo del modelo asociado a un objeto, sistema o fenómeno, independientemente del ámbito científico o aplicado en el cual tiene existencia, busca la caracterización cualitativa y cuantitativa de un aspecto específico del objeto del mundo real, generalmente dinámico (comportamientos, evolución y cambios de estado del sistema relacionado), de interés para la consecución de los objetivos del estudio basado en M&S.

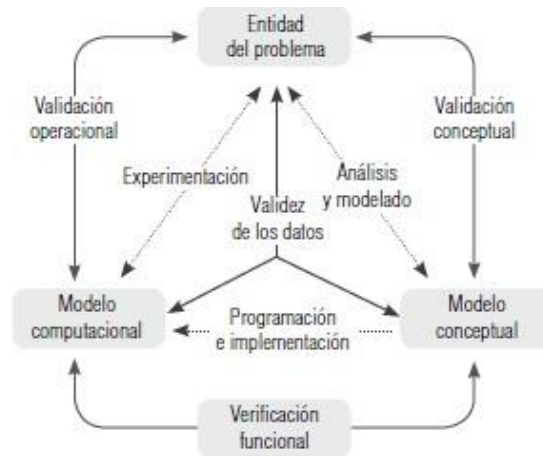
En el modelo para la simulación se pueden incluir, entre otras, las siguientes características:

- Relaciones matemáticas emergentes (alto nivel) entre determinados aspectos del sistema.
- Mecanismos subyacentes relativos a las relaciones entre los diferentes elementos del sistema.
- Los valores umbrales de los parámetros que determinan la dinámica del sistema a partir de cambios de estado.
- Estructuras y patrones característicos, coherentes con la formulación conceptual del modelo, en los resultados.
- Patrones de interacción y competencia entre estructuras coherentes.

### **2.3.3. *El proceso del M&S***

Desde la perspectiva de la NSF, citada anteriormente, el proceso de M&S, como componente del conocimiento científico, exige un proceso sistemático, riguroso y reproducible, que permita garantizar la calidad de sus resultados, tanto desde el punto de vista conceptual (cualitativo) en la caracterización del objeto de estudio como desde el

punto de vista de comportamiento (cuantitativo) en la implementación de los métodos y herramientas computacionales que permiten efectuar la experimentación numérica. Sargent (2010) Una de las principales barreras que enfrentará un programa de maestría en las áreas de mayor interés, es la diferencia del precio que el cliente potencial estima pagar por materia y los precios ofertados en la actualidad por la Universidad y sus competidores. presenta una versión sintetizada de los procesos y actividades requeridos en un estudio basado en modelado y simulación (ver figura 2-5).



**Figura 3-2.** Proceso de M&S en función de la validez del modelo obtenido

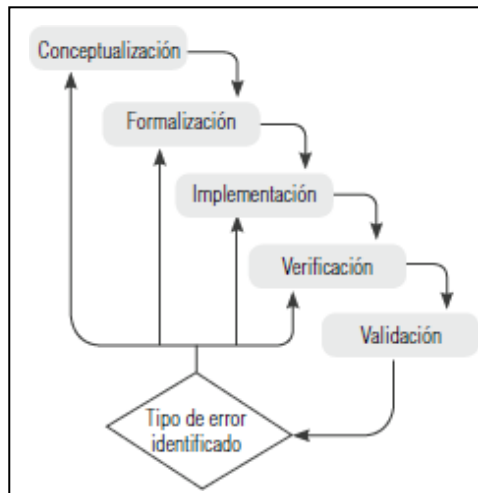
**Fuente:** Adaptación propia del gráfico presentado en Sargent, 2010.

Si bien en la mayoría de las áreas de aplicación el peso del trabajo relativo al modelado se centra en el desarrollo de un modelo para reproducir un comportamiento observado (y medido), en el M&S, como disciplina en sí misma, buena parte del trabajo se lleva a cabo en el componente de modelado en la búsqueda de dar respuesta a la pregunta de investigación que determina los objetivos del estudio.

En general, la aplicación del M&S a la solución de problemas específicos en ciencias básicas, ciencias sociales y económicas e ingeniería, en términos de objetivos como la caracterización, predicción, diseño, optimización, etc., se puede entender como un ciclo (que está, implícitamente, inspirado en el concepto, proveniente de la informática, del ciclo de vida del software) en el que intervienen de manera más o menos secuencial las etapas de conceptualización basada en descripciones en lenguaje natural, formalización en lenguaje matemático, implementación computacional, verificación funcional del modelo informático obtenido y validación general del modelo en función de los resultados obtenidos, tanto en los aspectos conceptuales como en los resultados numéricos consecuentes.

A continuación, se describe brevemente cada una de las etapas mencionadas:

- **Conceptualización (modelo narrativo):** análisis general del sistema real. Implica la determinación formal de los elementos, aspectos y relaciones relevantes en el sistema bajo estudio: definición de las partes relevantes, de los procesos y de las relaciones y un planteamiento claro del problema que se pretende solucionar.
- **Formalización (modelo esquemático):** definición de las variables del modelo. Selección y exclusión de partes y relaciones; escalas temporal y espacial para la solución del problema.
- **Implementación (modelo informático):** traducción del modelo esquemático a la representación codificada (en lenguaje de computador); reorganización de ecuaciones; representación apropiada de operaciones (estructuración del código y optimización).
- **Verificación funcional:** análisis de estabilidad. ¿Produce el modelo resultados razonables/ creíbles? (experiencia). Análisis de sensibilidad: variación de los resultados ante cambios en las variables dentro del rango de variación natural. Un parámetro crítico es aquel que induce cambios fuertes con pequeñas variaciones. Se deben analizar los resultados ante cambios en los parámetros dentro del error propio de cada uno.
- **Validación:** comprobación del modelo con datos independientes (fuentes de terceros). Comparación de los resultados del modelo respecto a modelos equivalentes en función (trabajos previos). La validación nunca es absoluta. Buenos resultados no garantizan un comportamiento correcto en todos los escenarios.



**Figura 4-2.** Ciclo del modelado y simulación

**Fuente:** Adaptación propia del gráfico presentado en Sargent, 2010.

Aquí se discriminan, según casos, los errores de mayor jerarquía para refinar, de manera cíclica, el modelo resultante y su implementación computacional con fines de experimentación numérica.

## CAPÍTULO III

### 3. METODOLOGÍA DE INVESTIGACIÓN

#### 3.1 Tipo y diseño de la investigación

Es importante manifestar que en esta investigación se la realizara de la manera descriptiva y correlacional.

Se lo hará de la manera descriptiva al tener un hecho que ha provocado un problema o fenómeno mundial, considerando que en la actualidad se tiene algunos modelos matemáticos que han ayudado a entender la dinámica de la infección por COVID-19. La diferencia en el manejo del virus entre los diferentes países hace necesario tener en cuenta que la pandemia se propaga por diferentes medios, es por eso que nos planteamos la necesidad de realizar un modelo que nos permita relacionar los casos Covi y los contaminantes ambientales, así se realizará un ajuste específico de los datos para nuestro escenario.

La presente investigación es de tipo descriptiva comparativa y correlacional (Sánchez y Reyes, 1996; Alarcón, 1991), es esencial manifestar que al manejar la investigación descriptiva nos ayudara a realizar una comparación para la recolección de datos con el propósito de analizar el comportamiento de una variable la cual puede afectar a una variable dependiente. En este enfoque de la investigación esta recolección de datos nos servirá para tener una visión más clara de la estructura con el firme propósito de obtener un módulo que permita representar la situación.

La importancia de esta investigación será para describir una correlación entre casos COVID y los contaminantes ambientales, para esto se genera una base de datos para facilitar el manejo de estos, considerando que la presente investigación correlacional va dirigida a la determinación del grado de relación existente entre dos variables o dos fenómenos o eventos suscitados.

En virtud, la investigación correlacional permitirá identificar las probables relaciones entre variables (Variable número de contagios COVID - 19) (Variable de contaminantes ambientales). Con este estudio llegaremos a obtener nuestro objetivo el mismo que es de brindar una explicación adecuado sobre el uso de modelos matemáticos

en pandemia específicamente SARS-CoV-2/COVID-19, usando las variables epidemiológicas como base de nuestro estudio. Además, Es considerable manifestar que el comportamiento de las epidemias influye determinadamente en la vida de la comunidad, al usar y entender el comportamiento de un virus en ciertas condiciones, las mismas que pueden ser usadas para estimar un escenario futuro de la pandemia

### **3.2 Métodos de investigación**

En esta investigación se trata sobre realidades de hechos y sus característica fundamental es buscar una interpretación correcta de los datos que se van a utilizar, es primordial fundamentar que tipo de análisis se utilizaran, para este caso de estudio se empezara con un método cuantitativo de los datos recopilados del número de contagios COVID registrados en la página de SALUD PUBLICA y de la página de monitoreo de la calidad de aire de la ciudad de QUITO, iniciando con un análisis descriptivo, luego un análisis inferencial y por último un análisis multivariante. «La investigación de tipo descriptiva trabaja sobre realidades de hechos, y su característica fundamental es la de presentar una interpretación correcta.

Para la investigación descriptiva, su preocupación primordial radica en descubrir algunas características fundamentales de conjuntos homogéneos de fenómenos, utilizando criterios sistemáticos que permitan poner de manifiesto su estructura o comportamiento. De esta forma se pueden obtener las notas que caracterizan a la realidad estudiada». (Sabino, 1986, pág. 51).

En esta investigación se utilizaron dos métodos de investigación: método deductivo y el método analítico.

Es importante manifestar que al utilizar el método deductivo la investigación se basa en hechos reales ocurridos en la actualidad de esto se hace necesario sacar conclusiones específicas. Según Bernal (2006) el método deductivo consiste en “tomar conclusiones generales para explicaciones particulares. El método se inicia con el análisis de los teoremas, leyes, postulados y principios de aplicación universal y de comprobada validez, para aplicarlos a soluciones o hechos particulares” (Bernal, 2006, pág. 56). En ese sentido Muños (2015), plantea lo siguiente: “Es el proceso lógico mediante el cual se parte de aspectos generales, aceptados como válidos, y los mismos, aplicando el razonamiento lógico, se puede explicar hechos particulares (Muños, 2015)

No obstante, no menos importante el método analítico nos ayudara a predecir un evento. Gómez (2012), el método analítico, “es un método de investigación, que consiste en descomponer el todo en sus partes, con el único fin de observar la naturaleza y los efectos del fenómeno. Sin duda, este método puede explicar y comprender mejor el fenómeno de estudio, además de establecer nuevas teorías.” (Gómez S. , 2012, pág. 15)

### **3.3 Enfoque de la investigación**

En esta investigación se aplicara un método cualitativo y cuantitativo debemos manifestar que implicara combinar los dos métodos en este estudio nos proporcionara resultados sobre algunas de sus variables y efectos sobre una determinada población en este caso la ciudad de QUITO, entonces se puede utilizar un estudio cualitativo en esta población para comprender mejor el fenómeno, en esta investigación vamos a verificar la naturaleza de la realidad del comportamiento del VIRUS su estructura dinámica, de esta manera determinar la fuerza asociada entre la relación de las dos variables, en general y como objeto los resultados de la investigación cualitativa conocida también como naturalista fenomenológica, interpretativa o etnográfica se enmarca dentro de un proceso que consta de las siguientes fases según (Hernández, Fernández & Baptista, 2010):

- Planteamiento del problema: El problema a resolver se plantea de forma general y amplia. Se caracteriza por la orientación hacia la exploración, la descripción y el entendimiento y está dirigido a las experiencias de los participantes.
- Revisión de la literatura: Tiene un rol secundario y se concibe como la justificación para el planteamiento y la necesidad del estudio.
- Recolección de datos: Los datos emergen poco a poco, se expresan en textos o imágenes y requieren un número relativamente pequeño de datos.
- Análisis de los datos: Se refiere al análisis de textos y material audiovisual, descripción análisis y desarrollo de temas y el significado profundo de los resultados
- Reporte de resultados: Debe ser emergente y flexible, reflexivo y con aceptación de tendencias

### **3.4 Alcance de la investigación**

El alcance de la investigación será de acuerdo con la información en los objetivos planteados.

¿Qué haré con el tema?

Como es conocido el caso que ocurrió a nivel mundial afectó la economía de muchos países en subdesarrollo, haciendo que su sistema de salud colapse de tal manera que no se sepa cómo combatir un virus invisible, por eso se debe hacer énfasis en la detección oportuna e inmediata y lo más importante el análisis de comportamiento del virus bajo ciertas variables, condiciones.

¿Qué haré con el problema de investigación?, ¿Qué uso le daré a la información recolectada?

Verificar y determinar las condiciones en las que el virus se comporta de manera agresiva, teniendo en cuenta que el COVID-19 muta de acuerdo con ciertas condiciones, la información manejará y permitirá contrarrestar en cierta medida la proliferación de más infectados y posterior deceso de los pacientes contagiados.

Para esta investigación se identifica la situación que se plantea con la finalidad de brindar la información necesaria para responder la problemática de la investigación. La investigación se centrará en el número de casos COVID 19, con los contaminantes en la ciudad de QUITO, se debe manifestar que, por la naturaleza de su comportamiento, en diferentes aspectos como su mutación y contagios, es importante describir que los datos deben satisfacer los parámetros requeridos para la confiabilidad, de esta manera la investigación maneja las características necesarias de las variables en el estudio.

### **3.5 Recolección de datos para la evaluación del estudio**

#### **3.5.1 Población**

La recolección de datos es una actividad en donde se deben agrupar y se recogen elementos importantes dentro de un contenido específico. Esto se hacen con el fin de generar información importante para el desarrollo de un trabajo determinada o un proyecto.

Para nuestro caso de estudio se utilizará el método de observación en vista que es una de las técnicas empleadas en las metodologías para la recolección de datos cualitativos. En donde se observa el objeto que será estudiado con el fin de que se determinen sus características.

Para ello, se debe analizar las particularidades de la situación por medio de observación. Al mismo tiempo se deben describir los elementos que se encuentran



integrados y que son parte del objeto de estudio. Los datos para utilizar fueron descargados de la página del Ministerio De Salud para el número de casos COVID, y los datos de los contaminantes de la (Secretaría de Ambiente del Distrito Metropolitano de Quito, (2022).

### **3.6 Datos**

La fuente de los datos que se utilizará para la aplicación de regresión lineal es la base de datos que se encuentra en la página de MINISTERIO DE SALUD PUBLICA, la página de contaminantes de QUITO. Es importante manifestar que los valores a considerar serán, los casos confirmados diarios y casos confirmados nuevos, PM2.5, O3, SO2, CO, PM10.

Si bien es cierto que al aplicar un modelo este debe ajustarse a la realidad por es necesario tener en cuenta las relaciones matemáticas al realizar una predicción con datos obtenidos de las situaciones planteadas.

#### **3.6.1 La regresión lineal**

Es una técnica de modelado estadístico que se emplea para describir una variable de respuesta continua como una función de una o varias variables predictoras. Puede ayudar a comprender y predecir el comportamiento de sistemas complejos o a analizar datos experimentales, financieros y biológicos. Las técnicas de regresión lineal permiten crear un modelo lineal. Este modelo describe la relación entre una variable dependiente y (también conocida como la respuesta) como una función de una o varias variables independientes  $X_i$  (denominadas predictores).

La ecuación general correspondiente a un modelo de regresión lineal es:

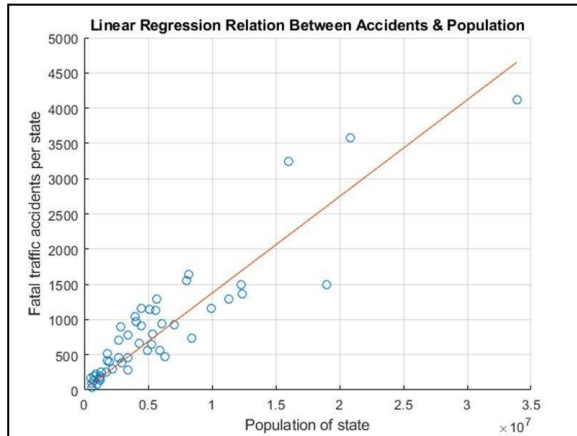
$$Y = Q_0 + \sum Q_i X_i + \epsilon_i \quad (32)$$

dónde  $\beta$  representa las estimaciones de parámetros lineales que se deben calcular y  $\epsilon$  representa los términos de error.

#### **3.6.2 Regresión lineal simple**

Son modelos que utilizan un único predictor. La ecuación general es:

$$Y = Q_0 + Q_0X + \epsilon_i \tag{33}$$



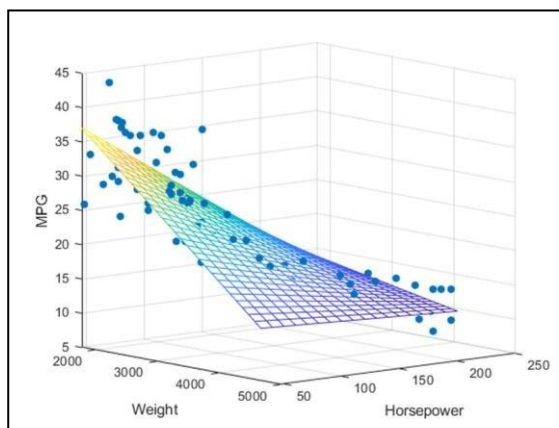
**Figura 5-3.** Ejemplo de regresión lineal simple

Fuente: MATLAB, 2022

### 3.6.2 Regresión lineal múltiple

Son los modelos que utilizan múltiples predictores. Esta regresión tiene múltiples  $X_i$  para predecir la respuesta,  $Y$ . Este es un ejemplo de la ecuación:

$$Y = Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \epsilon$$



**Figura 6-3** Regresión lineal múltiple, millas por galón (MPG)

Fuente: MATLAB, 2022

### 3.6.3 Regresión lineal multivariante

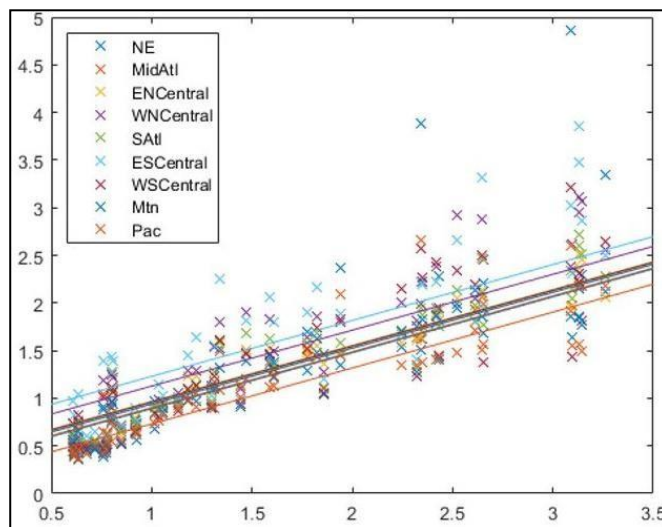
Son modelos para varias variables de respuesta. Esta regresión tiene múltiples  $Y_i$  que derivan de los mismos datos  $Y$ . Se expresan con fórmulas diferentes. Este es un ejemplo del sistema con 2 ecuaciones:

$$Y_1 = Q_{01} + Q_{11}X_1 + \epsilon_1$$

$$Y_2 = Q_{02} + Q_{12}X_1 + \epsilon_2$$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} Q_{01} & Q_{11} \\ Q_{02} & Q_{12} \end{pmatrix} \begin{pmatrix} \mathbf{1} \\ X_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

(34)



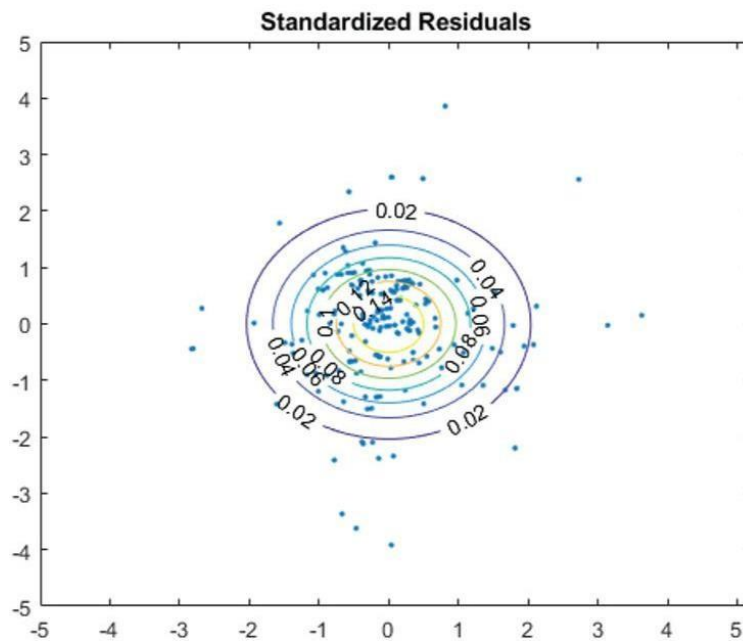
**Figura 7-3.** Ejemplo de regresión lineal multivariante

Fuente: MATLAB, 2022

### 3.6.4 Regresión lineal múltiple multivariante

Son modelos que utilizan varios predictores para múltiples variables de respuesta. Esta regresión tiene múltiples  $X_i$  para predecir varias respuestas  $Y_i$ . Esta es una generalización de las ecuaciones:

$$\begin{matrix}
 Y_1 & 1 & X_{11} & X_{12} & \dots & X_{1q} & Q_1 & \epsilon_1 \\
 Y_2 & 1 & X_{21} & X_{22} & \dots & X_{2q} & Q_2 & \epsilon_2 \\
 \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\
 Y_n & 1 & X_{n1} & X_{n2} & \dots & X_{nq} & Q_n & \epsilon_n
 \end{matrix}
 \quad (35)$$



**Figura 8-3.** Regresión lineal múltiple multivariante que calcula las MPG

Fuente: MATLAB, 2022

### 3.7 Selección de Muestra

Considerado que los datos se encuentran en la página de Ministerio de Salud Pública, además de la página de contaminantes de la ciudad de Quito. estos datos serán ordenadas en una base de datos, los mismos que serán utilizados para generación del modelo. Se debe tener en cuenta el análisis será desde enero 2020 – marzo 2021.

### 3.8 Tamaño de Muestra

Es importante manifestar que los datos a utilizarse será la población registrada en la página del Ministerio de Salud Pública, realizando una relación directa con los contaminantes.

Es importante manifestar que al comprobar los datos se hace muy necesario para validar nuestro caso de estudio, se aplicara dos métodos, para esto se debe mencionar que al tener una distribución dispersa se hace necesario aplicar.

Es importante mencionar que se debe aplicar pruebas no paramétricas. Se denominan pruebas no paramétricas aquellas que no presuponen una distribución de probabilidad para los datos que van a ser analizados, para ello se conocen como distribución libre (distribución free).

Los resultados estadísticos se derivan únicamente a partir de procedimientos de ordenación y recuento, por lo que la base lógica es de fácil comprensión. Cuando se trabaja con muestras pequeñas ( $n < 10$ ) en las que se desconoce si es válido suponer la normalidad de los datos, conviene en muchas ocasiones utilizar pruebas no paramétricas, al menos para corroborar resultados obtenidos a partir de la utilización de teoría basada en la normal.

En estos casos se emplea como parámetro de centralización la mediana, que es aquel punto para el que el valor de X está el 50% de las veces por debajo y el 50% por encima. Las pruebas que se aplicaran en nuestro caso de estudio van a hacer correlación de Pearson y Spearman, regresión lineal y múltiple. para los datos, por ello se conocen también como de distribución libre (distribución free). En la mayor parte de ellas los resultados estadísticos se derivan únicamente a partir de procedimientos de ordenación y recuento, por lo que su base lógica es de fácil comprensión.

### **3.9 Correlación de Spearman**

El coeficiente de correlación de Spearman es una prueba no paramétrica cuando se desea medir la relación entre dos variables y no se cumple el supuesto de normalidad en la distribución de tales valores. El coeficiente de correlación de Spearman se designa por  $r_s$ .

### **3.10 Pasos para la prueba del coeficiente de correlación de Spearman**

#### ***3.10.1 Formulación de hipótesis***

a) Prueba unilateral derecha  $H_0$ : X e Y son mutuamente independientes.  $H_1$ : Existe una tendencia a formar parejas entre los valores grandes de X e Y.

b) Prueba unilateral izquierda  $H_0$ : X e Y son mutuamente independientes.  $H_1$ : Existe una tendencia de los valores grandes de X a formar parejas con los valores pequeños de Y.

c) Prueba bilateral  $H_0$ : X e Y son mutuamente independientes.  $H_1$ : X e Y no son mutuamente independientes. Las hipótesis unilaterales indicadas en los incisos b) y c) se utilizan, respectivamente, cuando el investigador desea saber si es posible concluir que las variables están directa o inversamente relacionadas. Las hipótesis especificadas en el inciso a) conducen a una prueba bilateral, y se utilizan cuando se desea descubrir cualquier desviación de la independencia.

### 3.10.2 Nivel de significancia “ $\alpha$ ”

La estadística de prueba se determina con la siguiente ecuación:

$$r_s = 1 - \frac{\sum_{i=1}^n i^2}{n(n^2-1)} \quad (36)$$

Establecimiento de los criterios de decisión: Si  $n$  está entre 4 y 30, se compara el valor calculado de  $r_s$ , con los valores críticos, de la tabla P.

Para la prueba bilateral, se rechaza  $H_0$  en el nivel de significación  $\alpha$  si  $r_s$ , es mayor que o menor que está en la intersección de la columna encabezada por  $\alpha/2$  y el renglón que corresponde a  $n$ .

Para la prueba unilateral derecha que especifica una correlación directa, se rechaza  $H_0$  en el nivel de significación  $\alpha$  si  $r_s$  es mayor que para  $\alpha$  y  $n$ .

La hipótesis nula  $H_0$ , se rechaza en la prueba unilateral izquierda en el nivel de significación  $\alpha$  si  $r_s$  es menor que para  $\alpha$  y  $n$ . Opcionalmente también puede trabajar con la tabla Q para una prueba bilateral. Si  $n > 30$ , se puede calcular:

$$r_s = 1 - \frac{\sum_{i=1}^n i^2}{n(n^2-1)} \quad (37)$$

Y utilizar la tabla I y II para obtener los valores críticos.

En estadística, la prueba de Kruskal-Wallis (de William Kruskal y W. Allen Wallis) es un método no paramétrico para probar si un grupo de datos proviene de la misma población. Intuitivamente, es idéntico al ANOVA con los datos reemplazados por categorías. Es una extensión de la prueba de la U de Mann-Whitney para 3 o más grupos.

Ya que es una prueba no paramétrica, la prueba de Kruskal-Wallis no asume normalidad en los datos, en oposición al tradicional ANOVA. Sí asume, bajo la hipótesis nula, que los datos vienen de la misma distribución. Una forma común en que se viola este supuesto es con datos heterocedásticos. El método de determinación se realiza con la ecuación 38.

$$K = (N - 1) \frac{\sum_{i=1}^g n_i (r_i - r_i)^2}{\sum_{i=1}^g \sum_{j=1}^{n_i} (r_i - r_i)^2} \quad (38)$$

Dónde

$n_i$  es el número de observaciones en el grupo  $i$

$n_{ij}$  es el rango (entre todas las observaciones) de la observación  $j$ , en el grupo  $i$ .

$N$  es el número total de observaciones entre todos los grupos

$$r_i = 1 - \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i} \quad (39)$$

$r_i = (N + 1)/2$  es el promedio de  $r_{ij}$ .

Note que el denominar de la expresión para  $K$  es exactamente:

$$\frac{(N - 1)N(N + 1)}{12} \quad (40)$$

Se puede realizar una corrección para los valores repetidos dividiendo  $K$  por:

$$1 - \frac{\sum_i (t_i^3 - t_i)}{ni}$$

Dónde  $G$  es el número de grupos de diferentes rangos repetidos, y  $t_i$  es el número de observaciones repetidas dentro del grupo  $i$  que tiene observaciones repetidas para un determinado valor. Esta corrección hace cambiar a  $K$  muy poco al menos que existan un gran número de observaciones repetidas.

Finalmente, el  $p$ -value (valor  $p$ ) es aproximado por  $Pr(X_{g-1}^2 \geq K)$ . Si algún  $n_i$  es pequeño ( $>5$ ) la distribución de  $K$  puede ser distinta de la chi-cuadrado.



## CAPÍTULO IV

### 4. RESULTADOS Y DISCUSIÓN

#### 4.1 Generalidades

Se presentan los datos obtenidos del Ministerio de Salud Pública y de la Secretaría de Ambiente del Distrito Metropolitano de Quito de casos nuevos de Covid 19 y de los niveles de contaminantes atmosféricos (PM2.5, PM10, SO<sub>2</sub>, O<sub>3</sub> y CO) registrados en las estaciones de Centro histórico y Belisario.

Además, se realizó un análisis matemático de la relación de las variables de contagios vs contaminantes atmosféricos.

Considerando el comportamiento estadístico de los datos y aplicando pruebas de regresión lineal simple y múltiple, obteniendo los coeficientes de correlación de cada modelo.

##### 4.1.1 Contagios de Covid 19

Se recopilaron los datos de casos infectados diarios por Covid-19 en la ciudad de Quito-Ecuador en el periodo comprendido del 13 de marzo 2020 al 13 de marzo 2021 de la página del Ministerio de Salud Pública, se sistematizó los datos obteniendo un valor acumulado mensual de nuevos contagios para facilitar el análisis.

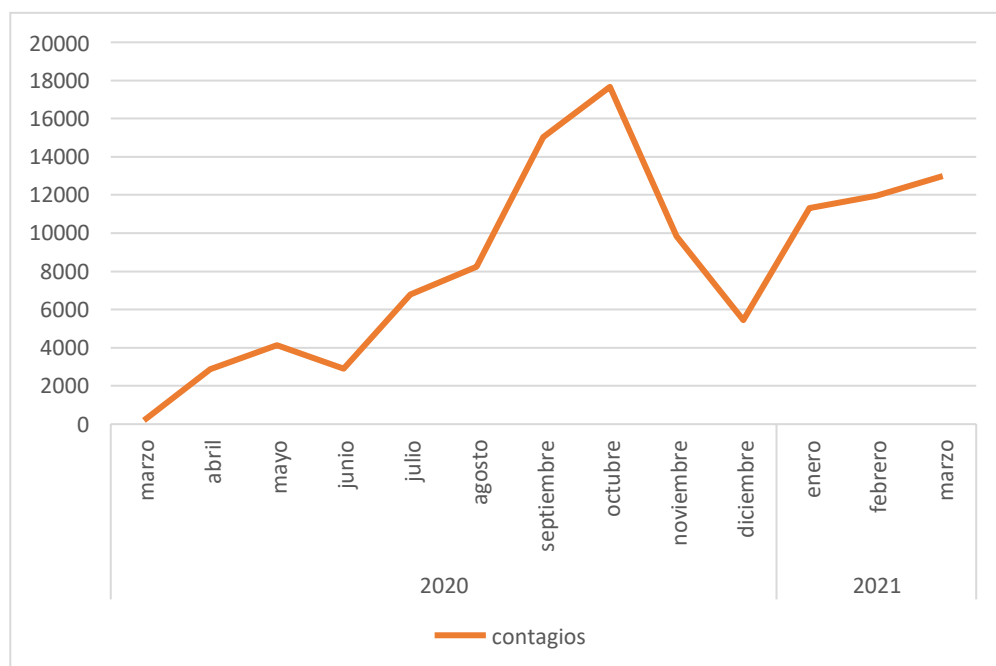
La tabla 4-4 muestran los contagios desde el mes de marzo de 2020 hasta el mes de marzo de 2021.

**Tabla 4-4** Contagios Covid 19

<b>Año</b>	<b>Meses</b>	<b>No. Contagios nuevos</b>
2020	Marzo	193
	Abril	2870
	Mayo	4129
	Junio	2904
	Julio	6771
	Agosto	8247
	Septiembre	15040
	Octubre	17662
	noviembre	9843
	diciembre	5449
2021	enero	11304
	febrero	11966
	marzo	12981

Fuente: Ministerio de Salud Pública (2022)

Realizado por: Pérez (2022)



**Figura 9-4.** Comportamiento de contagios de Covid-19 marzo 2020 – marzo 2021.

Realizado por: Pérez (2022)

Como se puede observar en la figura 9-4, los contagios de Covid-19 llegó a su valor más alto en el mes de octubre de 2020, con 18000 nuevos casos, seguido por el valor en marzo 2021 con 13000 nuevos casos.

Los valores más bajos registrados fueron en diciembre de 2020, y en el periodo marzo 2020 a junio 2020, sin embargo, en este último, los datos no se los considera reales debido a que el Ministerio de Salud Pública no mantenía un registro eficiente de casos nuevos.

#### 4.1. Contaminantes atmosféricos

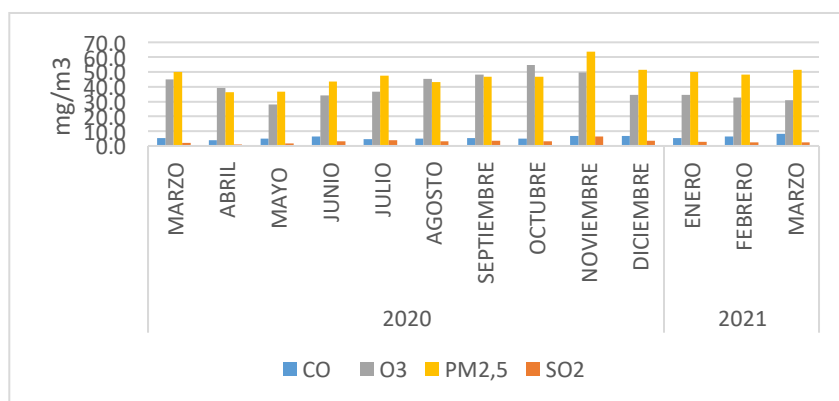
Se construyó la base de datos de los contaminantes del aire PM2,5, PM10, O3, SO2 y CO para la ciudad de Quito-Ecuador en el periodo comprendido del 13 de marzo 2020 al 13 de marzo 2021, mediante los valores obtenidos de dos estaciones de medición en la ciudad, ubicadas en el Centro histórico y en la parroquia Belisario.

**Tabla 5-4** Contaminantes

Meses		CO	O3	PM2,5	SO2
2020	Marzo	5,5	44,9	50,1	2,0
	Abril	4,1	39,0	36,5	1,0
	Mayo	5,0	27,9	36,7	1,9
	Junio	6,3	34,1	43,6	3,0
	Julio	4,8	36,7	47,6	3,8
	Agosto	5,1	45,4	43,1	3,3
	Septiembre	5,4	48,1	46,9	3,6
	Octubre	5,1	54,7	46,9	3,2
	Noviembre	6,7	49,6	63,8	6,6
	Diciembre	6,9	34,6	51,4	3,7
2021	Enero	5,5	34,6	49,9	2,7
	Febrero	6,3	32,9	48,1	2,4
	Marzo	8,1	31,1	51,5	2,5

Fuente: Secretaria de Ambiente del Distrito Metropolitano de Quito (2022)

Realizado por: Pérez (2022)



**Figura 10-4.** Comportamiento de los contaminantes atmosféricos

Realizado por: Pérez (2022)

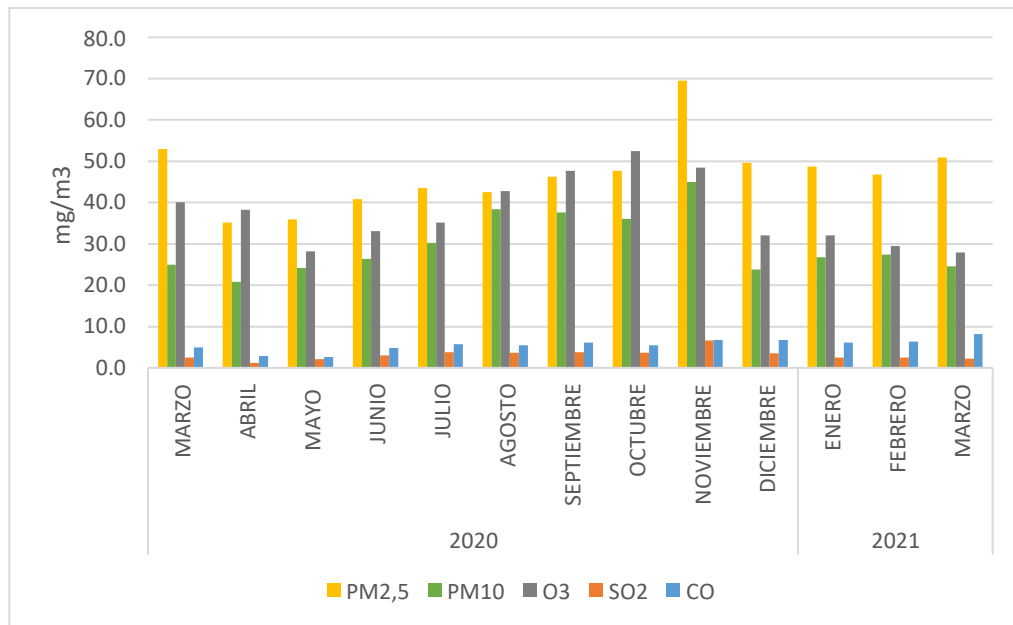
Se muestra que los parámetros de PM<sub>2,5</sub> y el O<sub>3</sub> son los que presentan mayores concentraciones en la atmosfera la parte céntrica de la ciudad, y los parámetros de Monóxido de Carbono (CO) y Dióxido de Azufre (SO<sub>2</sub>) presentan menor concentración. De forma general, se presenta los valores más altos de concentraciones en los meses de marzo de 2020, con 50 mg/m<sup>3</sup>, en noviembre de 2020 con 65 mg/m<sup>3</sup>. Los meses con valores más bajos se presentan en el mes de mayo de 2020.

**Tabla 6-4** Valores de concentración de contaminantes atmosféricos mg/m<sup>3</sup>

Meses	PM <sub>2,5</sub>	PM <sub>10</sub>	O <sub>3</sub>	SO <sub>2</sub>	CO	
2020	Marzo	53,0	25,0	40,0	2,5	5,0
	Abril	35,1	20,8	38,2	1,2	2,9
	Mayo	35,9	24,1	28,2	2,0	2,6
	Junio	40,8	26,4	33,1	3,0	4,8
	Julio	43,5	30,3	35,2	3,8	5,6
	Agosto	42,5	38,3	42,7	3,6	5,5
	Septiembre	46,3	37,6	47,7	3,7	6,0
	Octubre	47,6	36,1	52,5	3,6	5,5
	Noviembre	69,5	45,0	48,4	6,6	6,8
	Diciembre	49,6	23,8	32,0	3,5	6,7
2021	Enero	48,6	26,7	32,0	2,5	6,1
	Febrero	46,8	27,3	29,5	2,4	6,3
	Marzo	50,8	24,5	27,9	2,3	8,1

Fuente: Secretaria de Ambiente del Distrito Metropolitano de Quito (2022)

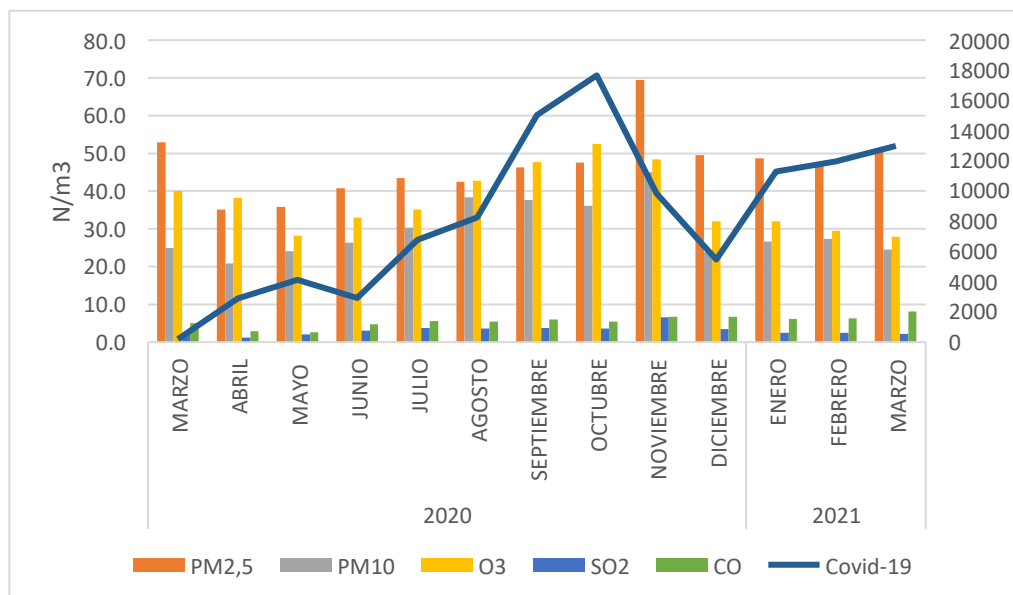
Realizado por: Pérez (2022)



**Figura 11-4.** Comportamiento de los contaminantes atmosféricos

Realizado por: Pérez (2022)

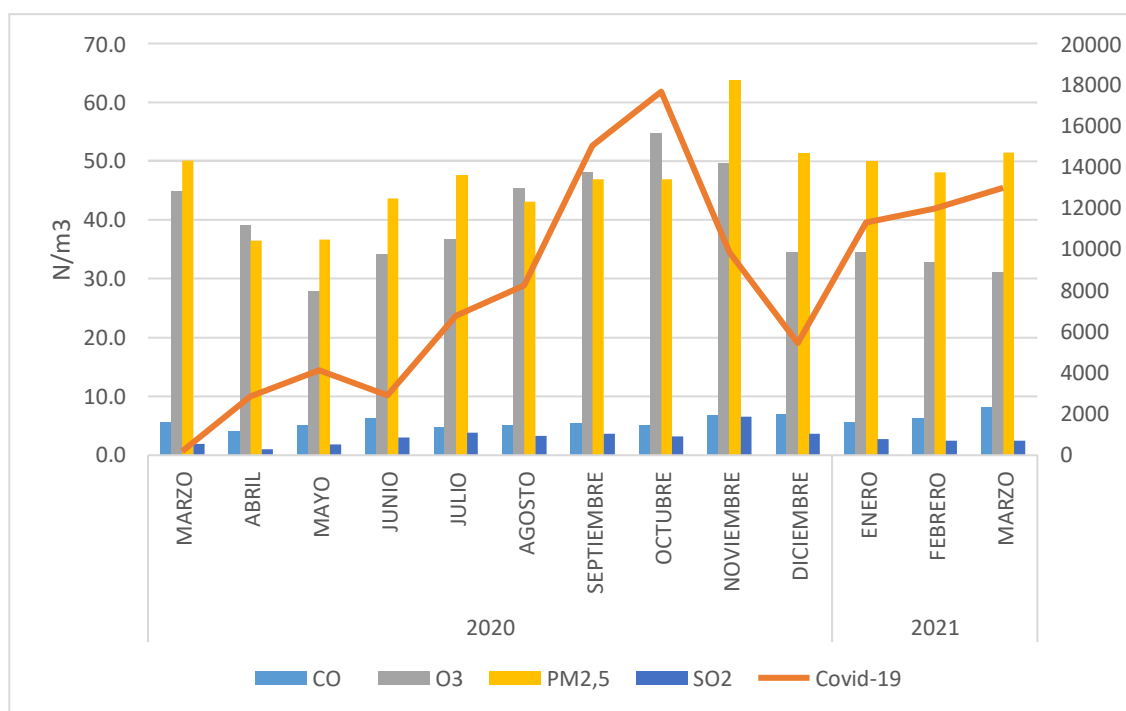
Como se observa en la figura 11-4, los niveles de PM<sub>2,5</sub>, ozono (O<sub>3</sub>) y PM<sub>10</sub> son los que presentan mayores concentraciones en la estación de Belisario con un pico de 70 mg/m<sup>3</sup>, 52 mg/m<sup>3</sup> y 45 mg/m<sup>3</sup> respectivamente. Los parámetros de SO<sub>2</sub> y monóxido de carbono (CO) presentan concentraciones bajas con relación al resto de parámetros.



**Figura 12-4.** Relación de los contaminantes atmosféricos y los contagios de Covid-19.

Realizado por: Pérez (2022)

La figura muestra la relación de las concentraciones de los contaminantes atmosféricos ( $PM_{2,5}$ ,  $PM_{10}$ ,  $O_3$ ,  $CO$  y  $SO_2$ ) en Belisario, con el número de casos nuevos, donde se observa que las concentraciones de los gases aumentan en el mes de noviembre de 2020, mientras los casos de Covid-19 mantuvieron su pico máximo en el mes de octubre de 2020. Por otro lado, en el mes de diciembre de 2020 los casos de Covid-19 bajaron, al igual que las concentraciones de los contaminantes.



**Figura 13-4.** Relación de los contaminantes atmosféricos y los contagios de covid

Realizado por: Pérez (2022)

La figura muestra la relación de las concentraciones de los contaminantes atmosféricos ( $PM_{2,5}$ ,  $PM_{10}$ ,  $O_3$ ,  $CO$  y  $SO_2$ ) en Centro Histórico, con el número de casos nuevos, donde se observa que las concentraciones de los gases aumentan en los meses de septiembre, octubre y noviembre de 2020, mientras los casos de Covid-19 mantuvieron su pico máximo en el mes de octubre de 2020. Por otro lado, en el mes de diciembre de 2020 los casos de Covid-19 bajaron, al igual que las concentraciones de los contaminantes.

## **4.2. Prueba de Normalidad**

Se estableció un análisis de normalidad de casos nuevos diarios por Covid-19 y el promedio de los valores de cada contaminante ( $PM_{2,5}$ ,  $PM_{10}$ ,  $O_3$ ,  $SO_2$  y  $CO$ ) de las estaciones de Belisario y Centro. Primeramente, se llevó a cabo un análisis de normalidad en SPSS según la metodología de Shapiro Wilks y de Kolmogorov-Smirnov.

**Tabla 7-4** Análisis de normalidad de los datos

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Y_Casos_Covid	0,116	64	0,033	0,917	64	0,000
X1_CO2	0,123	64	0,017	0,971	64	0,137
X2_O3	0,088	64	0,200*	0,982	64	0,498
X3_SO2	0,099	64	0,195	0,903	64	0,000
X4_CO	0,071	64	0,200*	0,976	64	0,256

*Nota.* \*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

**Realizado por:** Pérez (2022)

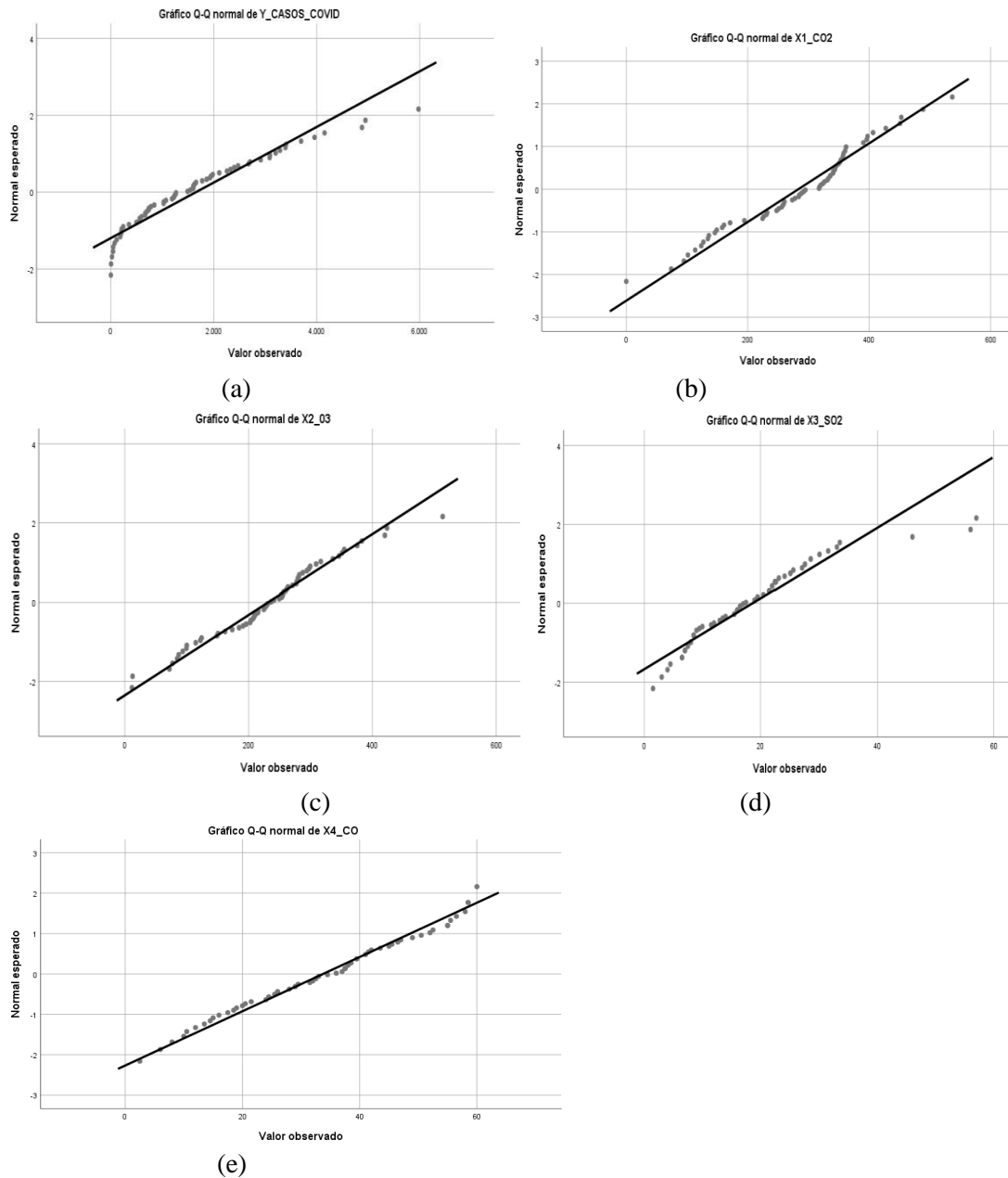
Se planteo la siguiente prueba de hipótesis con un nivel de significancia de 5% (0,05) y un nivel de confianza de 95%:

- Ho= Los datos de contagios de Covid-19 siguen una distribución normal
- H1=Los datos de contagios de Covid-19 no siguen una distribución normal

Debido a que el nivel de significancia de Shapiro Wilk es mayor que el coeficiente resultante en la variable Y y de X3 (SO<sub>2</sub>), se rechaza la hipótesis nula y se adopta la hipótesis alternativa, la cual menciona que los datos no siguen una distribución normal.

Por otro lado, el nivel de significancia mostrado en la tabla 4-4 es mayor que el nivel de significancia, por tanto, si presentan un comportamiento normal; esto se puede comprobar gráficamente mediante la figura 4-6, la misma que muestra que los datos de X1\_CO<sub>2</sub>, X2\_O<sub>3</sub>, y X4\_CO siguen un comportamiento normalizado.





**Figura 14-4.** Gráficos de pruebas de normalidad

Realizado por: Pérez (2022)

### 4.3. Correlación de Pearson y Spearman

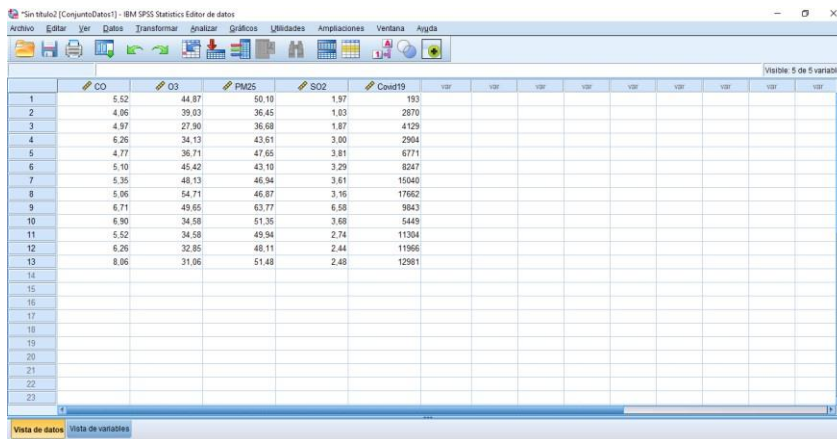
Se aplicaron dos metodologías de determinación de correlación entre las variables: la correlación de Pearson y Spearman en el software SPSS; esta prueba determina el grado de relación entre dos variables independientes, en este caso la relación entre los nuevos casos de Covid 19 y los contaminantes atmosféricos, tomando en consideración los rangos de la tabla 4-5.

**Tabla 8-4** Valores de la correlación

Rangos de correlación	Interpretación
0 – 0,25	Escasa o nula
0,26 – 0,5	Débil
0,51 – 0,75	Entre moderada y fuerte
0,76 – 1	Entre fuerte y perfecta

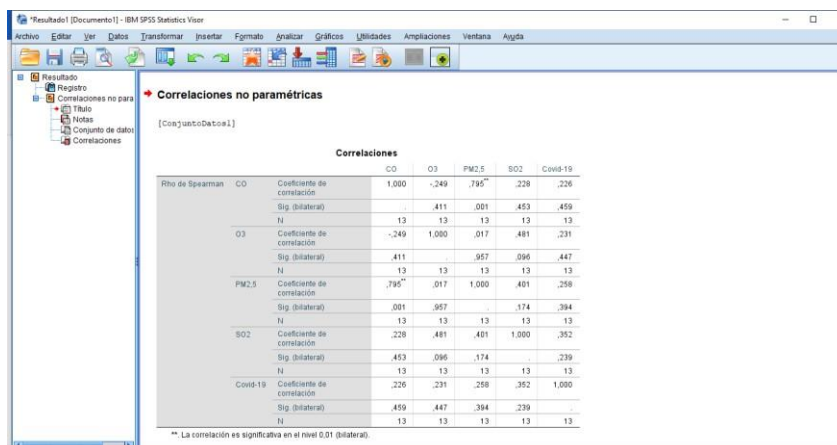
Realizado por: Pérez (2022)

Fuente: (Martínez R. , 2009)



**Figura 15-4.** Ingreso de datos al programa SPSS

Realizado por: Pérez (2022)



**Figura 16-4.** Salida de datos del programa SPSS

Realizado por: Pérez (2022)

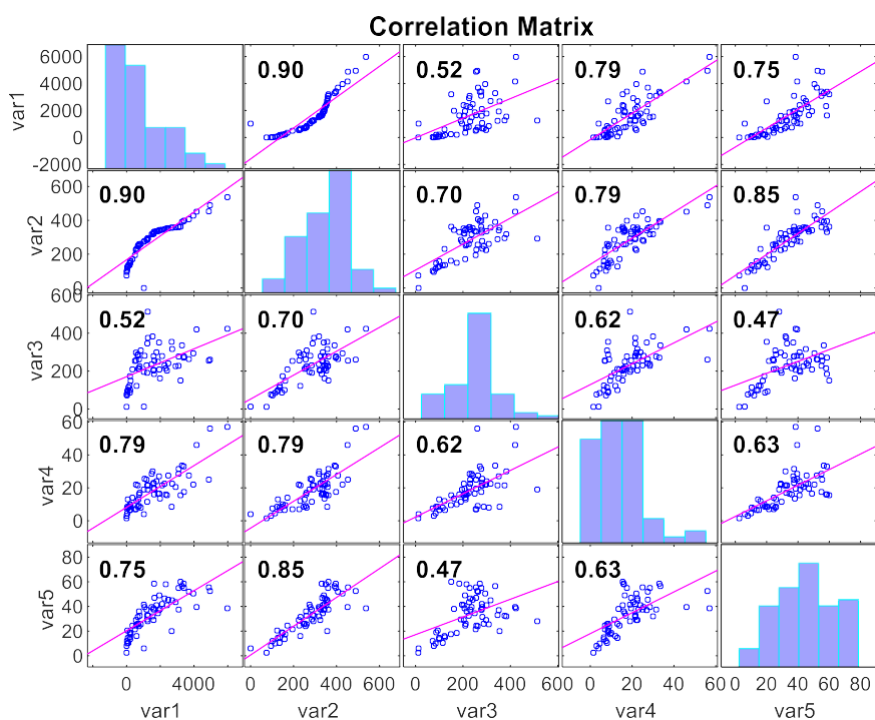
**Tabla 9-4** Valores de la correlación de los contaminantes con los casos de Covid-19

Contaminantes	Pearson	Spearman
X1_CO2	0,899**	0,985**
X2_O3	0,516**	0,552**
X3_SO2	0,785**	0,741**
X4_CO	0,755**	0,852**

*Nota.* \*\*. La correlación es significativa en el nivel 0,01 (bilateral).

Realizado por: Pérez (2022)

Como se observa en la tabla, existen correlaciones entre moderadas y altas de los datos analizados, obteniendo mejores resultados con la metodología de correlación de Spearman, donde se determinó una relación alta con 0,985 entre el CO<sub>2</sub> y los casos de Covid -19, y la más baja de 0,5, que se considera como una correlación moderada entre el ozono (O<sub>3</sub>) y los casos de Covid-19.



**Figura 17-4.** Análisis de correlación en Matlab

Realizado por: Pérez (2022)

En la figura 17-4 se observa que en referencia a la variable Y correspondiente a los casos de Covid-19 (var1), todas las variables presentaron una correlación alta, a excepción de var3, la cual arrojó una correlación moderada con respecto a la variable X2\_O<sub>3</sub> (var3), con un coeficiente de 0,52.

#### 4.4. Regresión lineal simple

Los análisis de regresión lineal se realizaron en el software SPSS y Matlab, para verificar la validez de los datos y contrastar los resultados obtenidos.

Se desarrollaron 4 modelos de regresión simple, los cuales se analizaron de la siguiente manera:

- Modelo 1: Contagios nuevos de Covid en el periodo marzo 2020 – marzo 2021 vs CO<sub>2</sub>.
- Modelo 2: Contagios nuevos de Covid en el periodo marzo 2020 – marzo 2021 vs O<sub>3</sub>.
- Modelo 3: Contagios nuevos de Covid en el periodo marzo 2020 – marzo 2021 vs SO<sub>2</sub>.
- Modelo 4: Contagios nuevos de Covid en el periodo marzo 2020 – marzo 2021 vs CO.

**Tabla 10-4** Valores de R de los modelos matemáticos

Modelo	SPSS				MATLAB	
	R	R cuadrado	R cuadrado ajustado	Error	R cuadrado	R cuadrado ajustado
X1_CO2	0,899 <sup>a</sup>	0,808	0,805	610,646	0,63795672	0,80798671
X2_O3	0,516 <sup>a</sup>	0,267	0,255	1193,443	0,26652424	0,26657474
X3_SO2	0,785 <sup>a</sup>	0,617	0,611	862,687	0,61316893	0,61677009
X4_CO	0,755 <sup>a</sup>	0,570	0,563	914,319	0,52636773	0,56952533

Realizado por: Pérez (2022)

#### 4.4.1. Modelo 1: Contagios nuevos de Covid vs CO<sub>2</sub>.

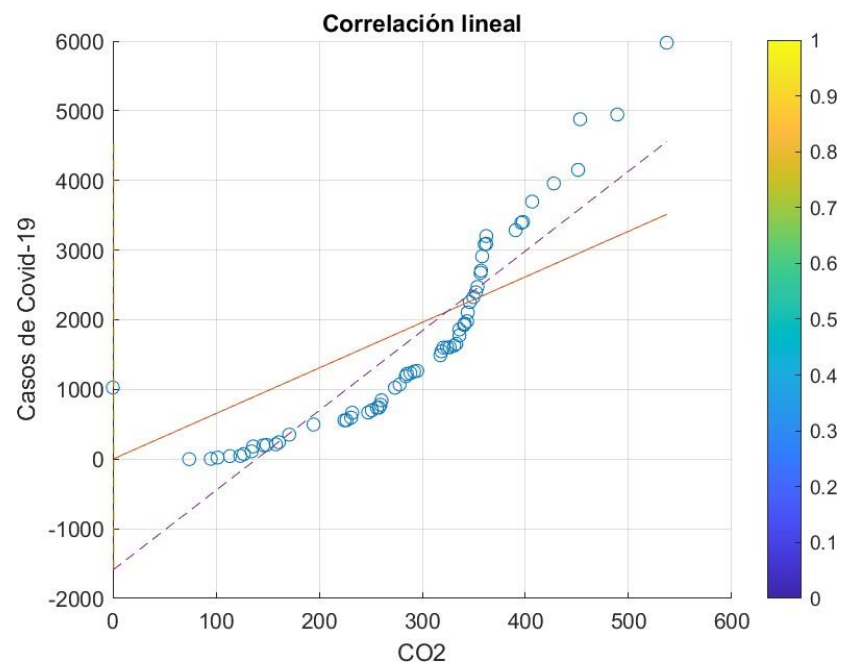
**Tabla 11-4** Coeficientes del modelo 1

Modelo		Coeficientes				
		Coeficientes no estandarizados		estandarizados		
		B	Desv. Error	Beta	t	Sig.
1	(Constante)	-1593,416	215,048		-7,410	0,000
	X1_CO <sub>2</sub>	11,454	0,709	0,899	16,152	0,000

**Realizado por:** Pérez (2022)

Según los coeficientes obtenidos del modelo 1 en el análisis de regresión lineal simple, se determina la siguiente ecuación:

$$y = 11,454x - 1593,416 \quad (41)$$



**Figura 18-4.** Regresión lineal de CO<sub>2</sub> vs contagios nuevos de Covid

**Realizado por:** Pérez (2022)

Los coeficientes de regresión lineal  $R^2$  y  $R^2$  ajustado obtenidos en SPSS son 0,808 y 0,805 respectivamente, y 0,6379 y 0,8079 en Matlab, observando que en este último se

obtuvo un valor ligeramente más elevado del modelo ajustado, denotando una correlación alta de las concentraciones de CO<sub>2</sub> y su influencia con los casos de Covid 19.

#### 4.4.2. Modelo 2: Contagios Covid en el periodo marzo 2020 – marzo 2021 vs O<sub>3</sub>.

**Tabla 12-4** Coeficientes del modelo 2

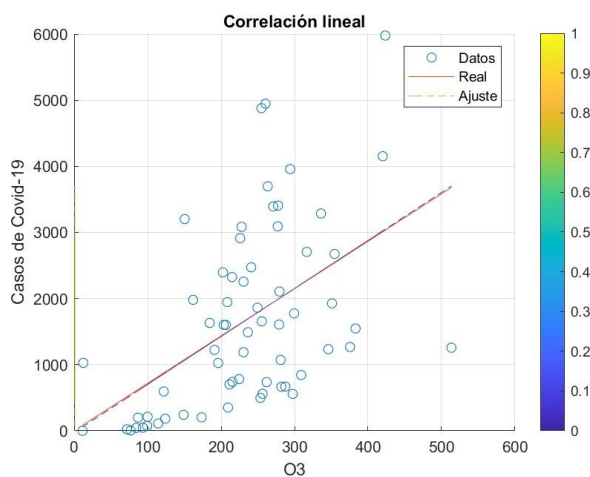
Modelo		Coeficientes				Sig.
		Coeficientes no estandarizados		estandarizados		
		B	Desv. Error	Beta	t	
2	(Constante)	-25,081	383,865		-0,065	0,948
	X2_O <sub>3</sub>	7,259	1,529	0,516	4,747	0,000

Realizado por: Pérez (2022)

Según los coeficientes obtenidos del modelo 2 en el análisis de regresión lineal simple, se determina la siguiente ecuación:

$$y = 7,259x - 25,081 \quad (42)$$

Cabe recalcar que el nivel de significancia de este modelo es mayor a 0,005, por lo que no es un modelo aceptable según los requerimientos del presente trabajo investigativo.



**Figura 19-4.** Regresión lineal de O<sub>3</sub> vs contagios nuevos de Covid

Realizado por: Pérez (2022)

Los coeficientes de regresión lineal  $R^2$  y  $R^2$  ajustado obtenidos en SPSS son 0,267 y 0,255 respectivamente, y 0,26652424 y 0,26657474 en Matlab, observando que en este último software se obtuvo un valor ligeramente más elevado del modelo ajustado, sin embargo, esta correlación es considerada como escasa según la tabla 4-5.

#### 4.4.3. Modelo 3: Contagios Covid en el periodo marzo 2020 – marzo 2021 vs $SO_2$ .

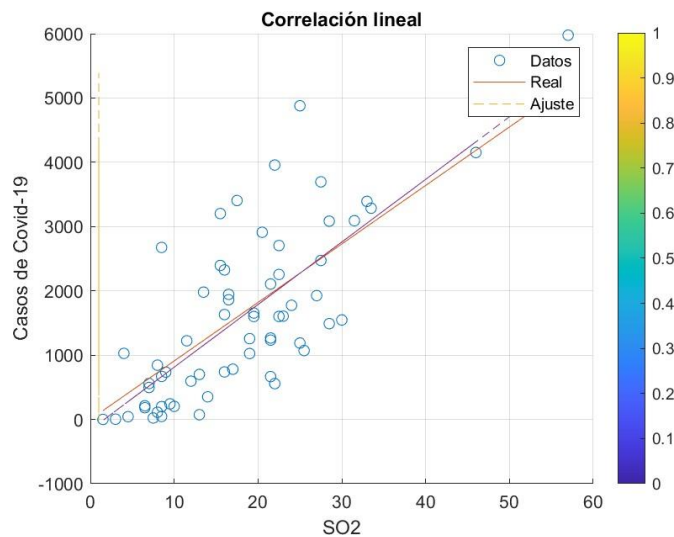
**Tabla 13-4** Coeficientes del modelo 3

Modelo		Coeficientes				
		Coeficientes no estandarizados		estandarizados		
		B	Desv. Error	Beta	t	Sig.
3	(Constante)	-161,287	211,306		-0,763	0,448
	X1_SO <sub>2</sub>	97,379	9,749	0,785	9,989	0,000

Realizado por: Pérez (2022)

Según los coeficientes obtenidos del modelo 3 en el análisis de regresión lineal simple, se determina la siguiente ecuación:

$$y = 97,379x - 161,287 \quad (43)$$



**Figura 20-4.** Regresión lineal de  $SO_2$  vs contagios nuevos de Covid

Realizado por: Pérez (2022)

En la figura 20-4 se puede observar la regresión lineal de los datos de  $SO_2$ , donde los coeficientes de regresión lineal  $R^2$  y  $R^2$  ajustado obtenidos en SPSS son 0,617 y 0,611

respectivamente, y 0,61316893 y 0,61677009 en Matlab, observando que en este último software se obtuvo un valor ligeramente más elevado del modelo ajustado, presentando una correlación moderada (tabla 4-5) entre el contaminante atmosférico SO<sub>2</sub> y los contagios de Covid-19

#### 4.4.4. Modelo 4: Contagios Covid en el periodo marzo 2020 – marzo 2021 vs CO.

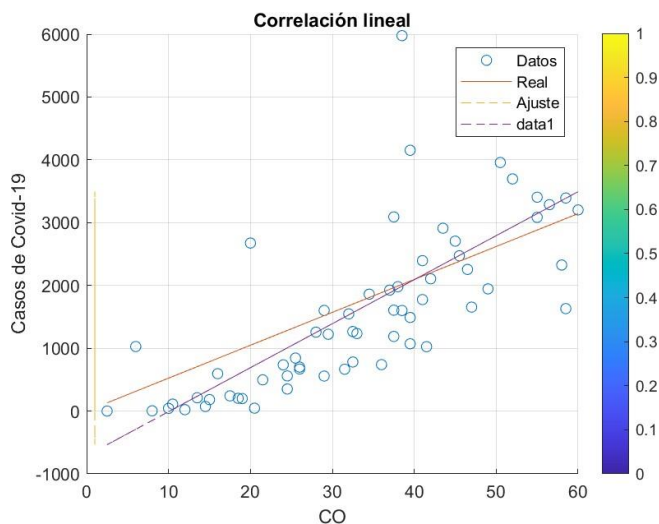
**Tabla 14-4** Coeficientes del modelo 4

Modelo		Coeficientes				Sig.
		Coeficientes no estandarizados		estandarizados		
		B	Desv. Error	Beta	t	
4	(Constante)	-710,518	284,987		-2,493	0,015
	X1_CO	70,057	7,735	0,755	9,057	0,000

Realizado por: Pérez (2022)

Según los coeficientes obtenidos del modelo 4 en el análisis de regresión lineal simple, se determina la siguiente ecuación:

$$y = 70,057x - 710,518 \quad (44)$$



**Figura 21-4.** Regresión lineal de CO vs contagios nuevos de Covid

Realizado por: Pérez (2022)

En la figura 21-4 se puede observar el análisis de regresión lineal de CO, y en la tabla se observa que esta variable presenta coeficientes de regresión lineal ajustado de



0,563 en obtenido en SPSS, y 0,56952533 obtenido en Matlab, adoptando el valor mayor correspondiente al análisis realizado en Matlab. Este valor de  $R^2$  denota una correlación moderada entre las variables CO y contagios de Covid-19.

#### 4.5. Regresión lineal múltiple

Los modelos desarrollados por regresión lineal múltiple fueron dos:

- Modelo 1: Contagios nuevos de Covid en el periodo marzo 2020 – marzo 2021 vs X1\_CO<sub>2</sub>, X2\_O<sub>3</sub>, X3\_SO<sub>2</sub>, X4\_CO.
- Modelo 2: Contagios nuevos de Covid en el periodo marzo 2020 – marzo 2021 vs X1\_CO<sub>2</sub>, X2\_O<sub>3</sub>, X3\_SO<sub>2</sub>.

##### 4.5.1. Modelo 1

La regresión lineal múltiple para el modelo 2 se realizó en el software SPSS y en Matlab, en donde se tomó como variable dependiente “Y” a los casos de Covid-19, y las variables independientes los cuales son los contaminantes atmosféricos X1\_CO<sub>2</sub>, X2\_O<sub>3</sub>, X3\_SO<sub>2</sub> y X4\_CO.

**Tabla 15-4** Resultados de los análisis de regresión lineal múltiple

	R	R2	R2 ajustado	Error estándar
SPSS	0,925 <sup>a</sup>	0,9	0,846	542
Matlab	0,825	0,856	0,846	542

Realizado por: Pérez (2022)

**Tabla 16-4** ANOVA aplicado al modelo matemático en SPSS

Modelo	Suma de cuadrados	Gl	Media cuadrática	F	Si g-
Regresión	103076	4	257690	87,	,00
	148,757		37,189	744	0 <sup>b</sup>
Residuo	173273	59	293684,		
	65,852		167		
Total	120403	63			
	514,609				

*Nota.* a. Variable dependiente: Y\_CASOS\_COVID

b. Predictores: (Constante), X4\_CO, X2\_O3, X3\_SO2, X1\_CO2

Realizado por: Pérez (2022)

Como se observa en la tabla, el modelo 1 de regresión lineal múltiple de las variables Y (Covid 19) y X1 (CO<sub>2</sub>), X2 (O<sub>3</sub>), X3 (SO<sub>2</sub>) y X4 (CO) arrojó un coeficiente de determinación (R<sup>2</sup>) en SPSS y en Matlab de 0,9 y 0,856 respectivamente, y el modelo ajustado tuvo un valor de 0,846 para ambos casos, el cual se ubica dentro del rango de correlaciones fuertes, validando el modelo matemático.

En la tabla de las pruebas de ANOVA se muestra el valor F, el cual se interpreta como la proporción de la mejora del modelo 1 con respecto a la inexactitud del mismo; se calculó un valor de 87,7, siendo mayor a 1, lo que se interpreta que el modelo es estadísticamente significativo.

**Tabla 17-4** Coeficientes beta del modelo matemático 1

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados	t	P	95,0% intervalo de confianza para B	
	B	Desv. Error				Beta	Límite inferior
(Constante)	-	205		-	0,	-	-
X1_CO	1255,479	,502	1,002	6,109	000	1666,687	844,271
X2_O3	,762	58		259	000	44	280
X3_SO2	-	1,0	-,272	-	0,	-	-
X4_CO	3,822	27	,236	3,721	000	5,877	1,767
(Constante)	29	10,		2,	0,	8,8	49,
X1_CO	,272	218		865	006	26	719
X2_O3	-	9,2	-,118	-	0,	-	7,4
X3_SO2	10,969	24		1,189	239	29,426	88

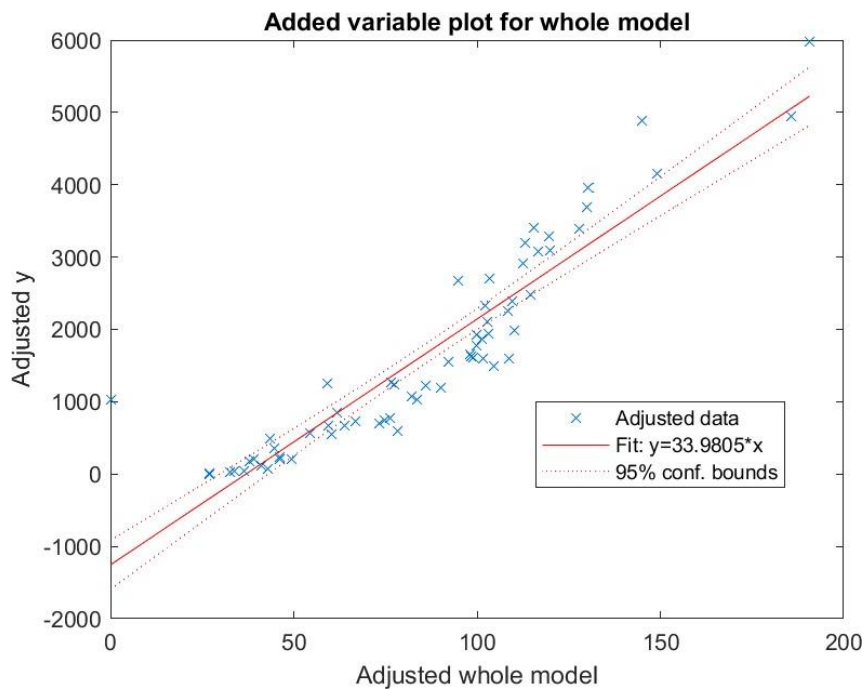
a. Variable dependiente: Y\_CASOS\_COVID

Realizado por: Pérez (2022)

Los valores “p” de las variables X1\_CO<sub>2</sub>, X2\_O<sub>3</sub> y X3\_SO<sub>2</sub> están por debajo del nivel de significancia de 0,05, por lo que se rechaza la hipótesis nula y se interpreta que estas variables realizan un aporte significativo al modelo. por otro lado, el valor de “p” de la variable X4\_CO es mayor que 0,05, por tanto, se acepta la hipótesis nula y se dice que el CO no realiza un aporte significativo al modelo; con este análisis se puede suprimir la variable X4\_CO y mejorar la exactitud del modelo matemático.

La ecuación que se obtuvo del modelo 1 con regresión lineal múltiple fue la siguiente:

$$y = 12,762X_1 - 3,822X_2 + 29,272X_3 - 10,969X_4 - 1255,479 \quad (45)$$



**Figura 22-4.** Regresión lineal múltiple del modelo 1

Realizado por: Pérez (2022)

#### 4.5.2. Modelo 2

La regresión lineal múltiple para el modelo 2 se realizó en el software SPSS y en Matlab, en donde se tomó como variable dependiente “Y” a los casos de Covid-19, y las variables independientes los cuales son los contaminantes atmosféricos X1\_CO<sub>2</sub>, X2\_O<sub>3</sub> y X3\_SO<sub>2</sub>.

**Tabla 18-4** Resultados de los análisis de regresión lineal múltiple del modelo 2

	R	R <sup>2</sup>	R <sup>2</sup> ajustado	Error estándar
SPSS	0,823 <sup>a</sup>	0,853	0,845	543,8
Matlab	0,825	0,853	0,845	544

Realizado por: Pérez (2022)

**Tabla 19-4** ANOVA aplicado al modelo matemático en SPSS

Modelo	Suma de cuadrados	gl	Media cuadrática	F	Si g.
Regresión	102660	3	342202	11	,00
	826,957		75,652	5,722	0 <sup>b</sup>
Residuo	177426	60	295711,		
	87,652		461		
Total	120403	63			
	514,609				

*Nota.* a. Variable dependiente: Y\_CASOS\_COVID

b. Predictores: (Constante), X4\_CO, X2\_03, X3\_SO2, X1\_CO2

Realizado por: Pérez (2022)

Como se observa en la tabla, el modelo 2 de regresión lineal múltiple de las variables Y (Covid 19) y X1 (CO<sub>2</sub>), X2 (O<sub>3</sub>) y X3 (SO<sub>2</sub>) arrojó un coeficiente de determinación (R<sup>2</sup>) en SPSS y en Matlab de 0,853 respectivamente, y el modelo ajustado tuvo un valor de 0,845 para ambos casos, el cual se ubica dentro del rango de correlaciones fuertes, validando el modelo matemático.

En la tabla de las pruebas de ANOVA se muestra el valor F, el cual se interpreta como la proporción de la mejora del modelo 2 con respecto a la inexactitud del mismo; se calculó un valor de 115,722, que con respecto al modelo 1, es mayor, es decir es más exacto.

**Tabla 20-4** Coeficientes beta del modelo matemático 2

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
	B	Desv. Error	Beta		
(Constante)	-1277,093	205,402		-6,218	,000
1 X1_CO2	11,169	1,142	,876	9,777	,000
X2_O3	-3,458	,984	-,246	-3,515	,001
X3_SO2	30,279	10,218	,244	2,963	,004

Nota. a. Variable dependiente: Y\_CASOS\_COVID

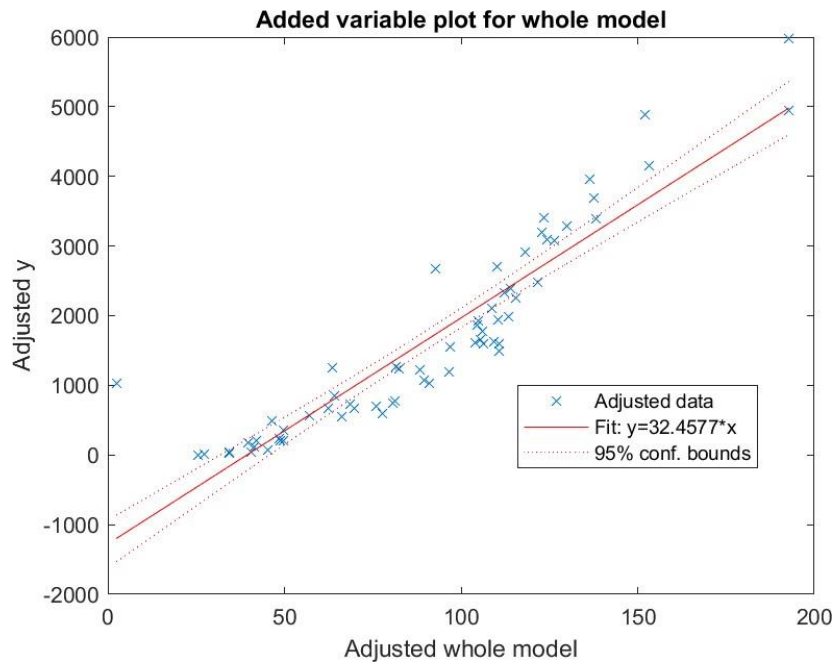
Realizado por: Pérez (2022)

#### 4.8 Decisión de prueba de hipótesis

Los valores “p” de las variables X1\_CO<sub>2</sub>, X2\_O<sub>3</sub> y X3\_SO<sub>2</sub> están por debajo del nivel de significancia de 0,05, por lo que se acepta la hipótesis nula y se interpreta que todas las variables independientes realizan un aporte significativo al modelo 2, mejorando su validez.

La ecuación que se obtuvo del modelo 2 con regresión lineal múltiple fue la siguiente:

$$y = 11,169X1 - 3,458X2 + 30,279X3 - 1277,093 \quad (46)$$



**Figura 23-4.** Regresión lineal múltiple del modelo 2

Realizado por: Pérez (2022)

## 4.9 Análisis comparativo en R

El análisis realizado también se lo realizó en R, el mismo que nos permitió realizar un el análisis que nos permita correlacionar cada variable independiente con la dependiente, de la misma manera se presentan los datos obtenidos del Ministerio de Salud Pública y de la Secretaria de Ambiente del Distrito Metropolitano de Quito de casos nuevos de Covid 19 y de los niveles de contaminantes atmosféricos. Se procede a realizar un análisis matemático de la relación de las variables de contagios vs contaminantes atmosféricos, considerando el comportamiento estadístico de los datos y aplicando pruebas de regresión lineal simple y múltiple, obteniendo los coeficientes de correlación de cada modelo.

### 4.9.1 Regresión Lineal

#### 4.9.1.1 Modelo 1: Contagio Covid vs X1\_CO2

```
Residuals:
  Min       1Q   Median       3Q      Max
-622.83 -484.03  -99.87   404.40 2620.42

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1593.4163   215.0484   -7.41 4.19e-10 ***
X1_CO2         11.4545     0.7092   16.15 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 610.6 on 62 degrees of freedom
Multiple R-squared:  0.808,    Adjusted R-squared:  0.8049
F-statistic: 260.9 on 1 and 62 DF,  p-value: < 2.2e-16
```

**Figura 25-4.** Regresión lineal simple Contagio Covid vs X1\_CO2

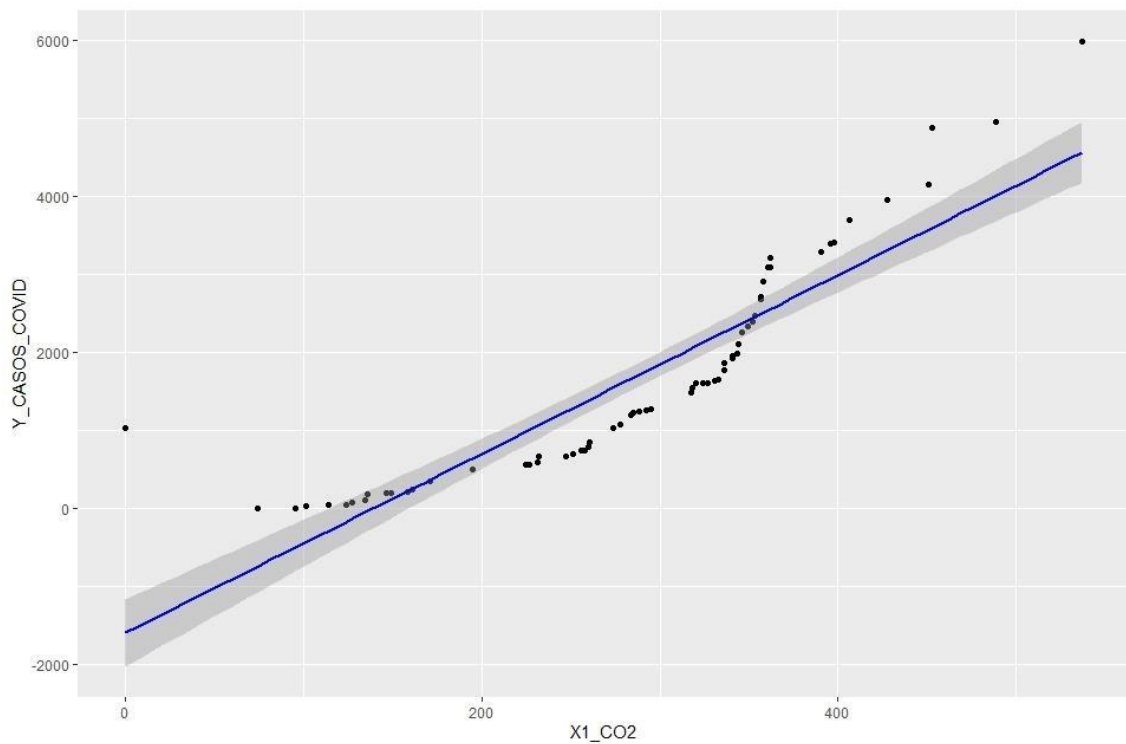
Realizado por: Pérez (2022)

Los resultados observados podemos observar los datos residuales y los interceptados los mismos que nos servirán para el valor Alfa y el valor beta, nuestro valor alfa o intercepto es de -1593.4163, nuestro  $\beta_1$  es 11.4545, estos valores nos muestran que los valores son significativos, por lo tanto el valor del coeficiente es aceptable, nuestro R cuadrado es de 0,808 y un R cuadrado ajustado de 0,8049 y al aplicar una correlación de Pearson es de 0,8988882 nuestro valor esta dentro del los limites establecidos en tablas la misma que esta en un intervalo de [-1,1], para nuestro caso existe una correlación

positiva, nuestro R cuadrado es confiable, es decir explica el comportamiento de las variables.

Según los coeficientes obtenidos del modelo 1 de regresión lineal con R en el análisis de regresión lineal simple, se determina la siguiente ecuación:

$$y = -1593.4163 + 11.4545X_1 \quad (47)$$



**Figura 26-4.** Regresión lineal simple Casos Covid – C02

Realizado por: Pérez (2022)

#### 4.9.1.2 Modelo 2: Contagio Covid vs X2\_03

Residuals:				
Min	1Q	Median	3Q	Max
-2446.2	-811.2	-319.0	799.4	3080.2

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-25.081	383.865	-0.065	0.948
X2_03	7.259	1.529	4.747	1.26e-05 ***

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1193 on 62 degrees of freedom  
 Multiple R-squared: 0.2666, Adjusted R-squared: 0.2547  
 F-statistic: 22.53 on 1 and 62 DF, p-value: 1.259e-05

**Figura 27-4.** Regresión lineal simple Casos Covid – X2\_03

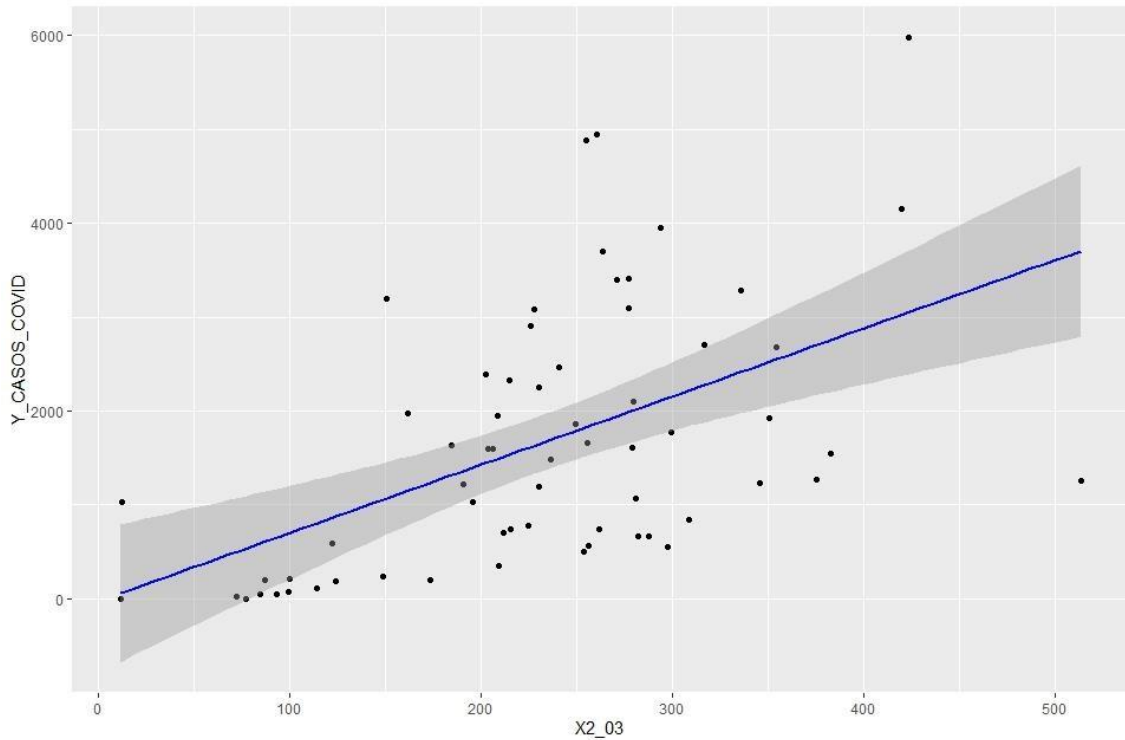
Realizado por: Pérez (2022)

En los resultados obtenidos podemos observar los datos residuales y los interceptados los mismos que nos servirán para el valor Alfa y el valor beta, nuestro valor alfa o intercepto es de -25.081, nuestro  $\beta_1$  es 7.259, estos valores nos muestran que el los valores son significativos, por lo tanto el valor del coeficiente es aceptable, nuestro R cuadrado es de 0,266 y un R cuadrado ajustado de 0,2547 y al aplicar una correlación de Pearson es de 0.5163332 nuestro valor es muy bajo a lo ideal, este valor debe acercarse a 1 o al 100%, para nuestro caso no existe una correlación, es un modelo muy bajo de confianza, nuestro R cuadrado no es confiable, es decir esta variable independiente no explica a mi variable dependiente no explica su comportamiento.

Según los coeficientes obtenidos del modelo 2 de regresión lineal con R en el análisis de regresión lineal simple, se determina la siguiente ecuación:

$$y = -25.081 + 7.259X_1 \quad (48)$$





**Figura 28-4.** Regresión lineal simple Casos Covid – X2\_03

Realizado por: Pérez (2022)

**4.9.1.3 Modelo 3: Contagio Covid vs X2\_SO2**

```

Residuals:
  Min       1Q   Median       3Q      Max
-1424.1  -543.2  -131.7   358.7  2605.8

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -161.287    211.306  -0.763    0.448
X3_SO2         97.379     9.749    9.989 1.56e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 862.7 on 62 degrees of freedom
Multiple R-squared:  0.6168,    Adjusted R-squared:  0.6106
F-statistic: 99.78 on 1 and 62 DF,  p-value: 1.556e-14

```

**Figura 29-4.** Regresión lineal simple Casos Covid – X3\_SO2

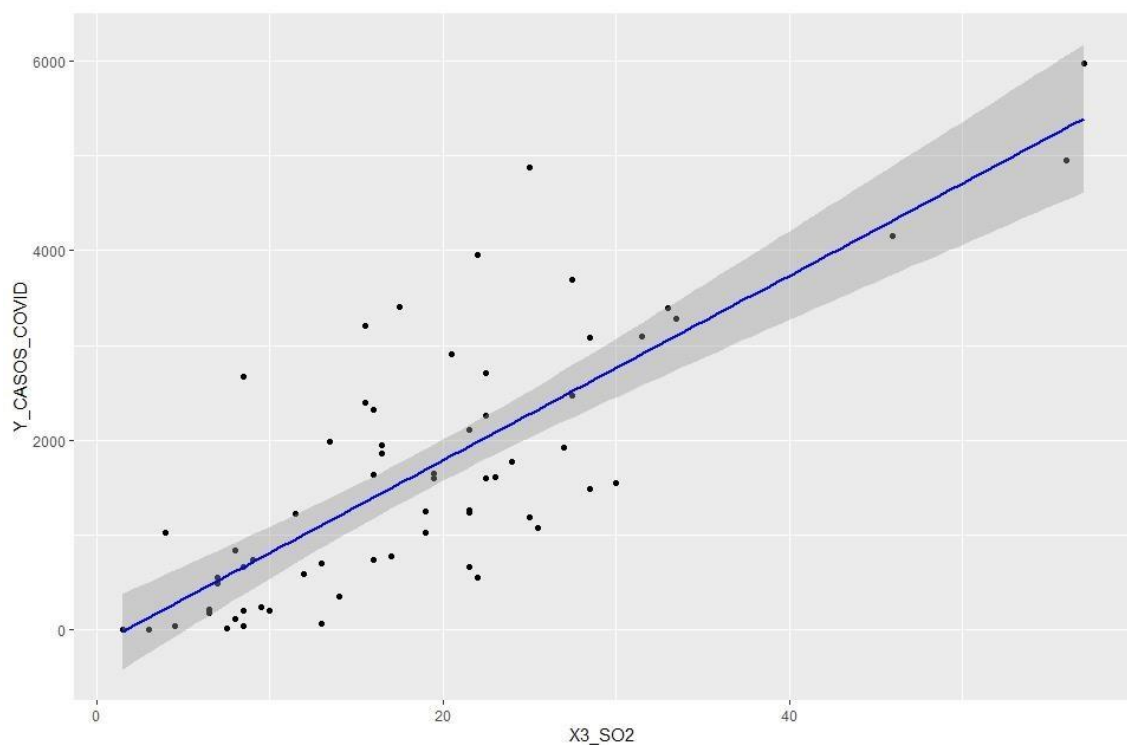
Realizado por: Pérez (2022)

En los resultados obtenidos podemos observar los datos residuales y los interceptados los mismos que nos servirán para el valor Alfa y el valor beta, nuestro valor alfa o intercepto es de -161.287, nuestro  $\beta_1$  es 97.379, estos valores nos muestran que el

los valores son significativos, por lo tanto el valor del coeficiente es aceptable, nuestro R cuadrado es de 0,6168 y un R cuadrado ajustado de 0.6106 y al aplicar una correlación de Pearson es de 0.7853662 nuestro valor es una correlación alta, este valor debe acercarse a 1 o al 100%, para nuestro caso existe una correlación alta, es un modelo moderado de confianza, nuestro R cuadrado es confiable, es decir esta variable independiente intenta explicar a mi variable dependiente no explica su comportamiento en su totalidad.

Según los coeficientes obtenidos del modelo 3 de regresión lineal con R en el análisis de regresión lineal simple, se determina la siguiente ecuación:

$$y = -161.287 + 97.379X_1 \quad (48)$$



**Figura 30-4.** Regresión lineal simple Casos Covid – X2\_SO2

**Realizado por:** Pérez (2022)

#### 4.9.1.4 Modelo 4: Contagio Covid vs X2\_CO

```

Residuals:
  Min       1Q   Median       3Q      Max
-1756.8  -426.7  -145.8   162.1  3990.3

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -710.518    284.987   -2.493   0.0153 *
X4_CO         70.057      7.735    9.057  5.94e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 914.3 on 62 degrees of freedom
Multiple R-squared:  0.5695,    Adjusted R-squared:  0.5626
F-statistic: 82.03 on 1 and 62 DF,  p-value: 5.938e-13

```

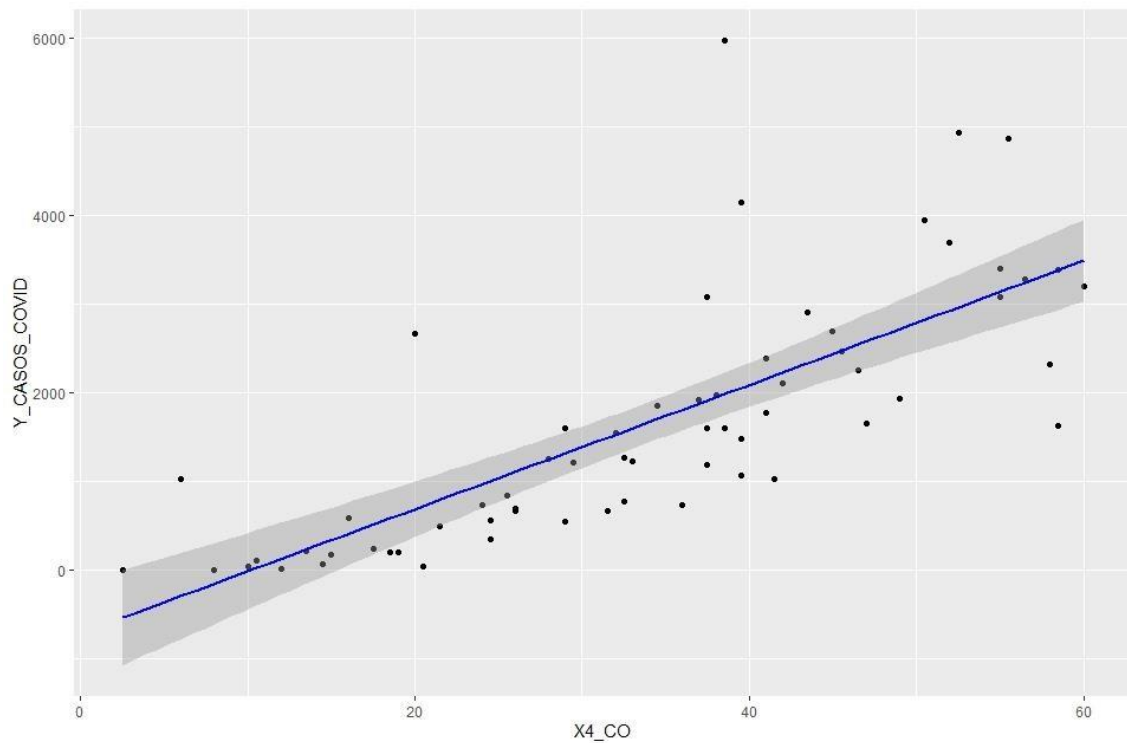
**Figura 31-4.** Regresión lineal simple Casos Covid – X3\_CO

Realizado por: Pérez (2022)

En los resultados obtenidos, podemos observar los datos residuales y los interceptados los mismos que nos servirán para el valor Alfa y el valor beta, nuestro valor alfa o intercepto es de -710.518, nuestro  $\beta_1$  es 70.057, estos valores nos muestran que son significativos, no tienden en gran medida a cero, por lo tanto el valor del coeficiente es aceptable, nuestro R cuadrado es de 0.5695 y un R cuadrado ajustado de 0.5626 y al aplicar una correlación de Pearson es de 0.7546522 nuestro valor es una correlación alta, este valor debe acercarse a 1 o al 100%, para nuestro caso existe una correlación alta, es un modelo moderado de confianza, nuestro R cuadrado es confiable, es decir esta variable independiente explica a mi variable dependiente.

Según los coeficientes obtenidos del modelo 4 de regresión lineal con R en el análisis de regresión lineal simple, se determina la siguiente ecuación:

$$y = -710.518 + 70.057X_1 \quad (48)$$



**Figura 32-4.** Regresión lineal simple Casos Covid – X2\_SO2

Realizado por: Pérez (2022)

#### 4.8.2 Mejor modelo

Al elegir el mejor modelo es importante manifestar que se desea incluir a mi modelación las variables que permiten especificar que afecta a la respuesta para así evitar resultados sesgados.

	Df	Sum of Sq	RSS	AIC
<none>			17742688	810.09
- X3_SO2	1	2596480	20339168	816.83
- X2_O3	1	3653668	21396356	820.07
- X1_CO2	1	28264389	46007076	869.07

**Figura 33-4.** Selección de mejor modelo

Realizado por: Pérez (2022)

En la imagen podemos apreciar que los mejores modelos analizados en regresión lineal son los modelos (1, 2, 3).

```

Residuals:
  Min       1Q   Median       3Q      Max
-824.26 -366.61  -66.54   281.95 2226.20

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1277.0928   205.4016  -6.218 5.31e-08 ***
X1_CO2       11.1689     1.1424   9.777 4.99e-14 ***
X2_O3        -3.4576     0.9837  -3.515 0.000843 ***
X3_SO2       30.2787    10.2183   2.963 0.004360 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 543.8 on 60 degrees of freedom
Multiple R-squared:  0.8526,    Adjusted R-squared:  0.8453
F-statistic: 115.7 on 3 and 60 DF,  p-value: < 2.2e-16

```

**Figura 34-4.** Selección de mejor modelo análisis de Estimados.

Realizado por: Pérez (2022)

Como datos obtenidos de los mejores modelos tomando en cuenta el R cuadrado previsto está en (0.8526), R cuadrado-ajustados (0.8453), el valor P de los mejores modelos es inferior a 0.05, su correlación es aceptable se acerca al 100%, se debe entender que cada vez que se añade un predictor, esto nos puede engañar de forma que se especifique un modelo demasiado complejo.

Como los mejores modelos de acuerdo con los datos obtenidos son:

- X1\_CO2
- X2\_O3
- X3\_SO2

#### 4.8.3 Prueba de normalidad

##### 4.8.3.1 Shapiro Test

**Tabla 21-4** SHAPIRO – WILK TEST DE NORMALIDAD

SHAPIRO - WILK TEST DE NORMALIDAD	
W =	0.91311
P - VALUE =	0.0002594

**Nota:** El valor obtenido en P-VALUE es de 0.0002594 RECHAZAMOS LA HIPOTESIS NULAS, es muy pequeño a 0.05 por lo tanto los valores son fiables y los resultados son importantes.

#### 4.8.3.2 Kolmogorov – Smirnov

Tabla 22-4 Kolmogorov – Smirnov test de normalidad

KOLMOGOROV – SMIRNOV TEST DE NORMALIDAD	
D =	0.11093
P -VALUE =	0.04876

Nota: El valor obtenido en P-VALUE es de 0.04876 NO SE RECHAZA LA HIPOTESIS NULA, el valor es mayor a 0.05 por lo tanto no se tiene evidencia necesaria para rechazar la HIPOTESIS NULA la distribución es normal.

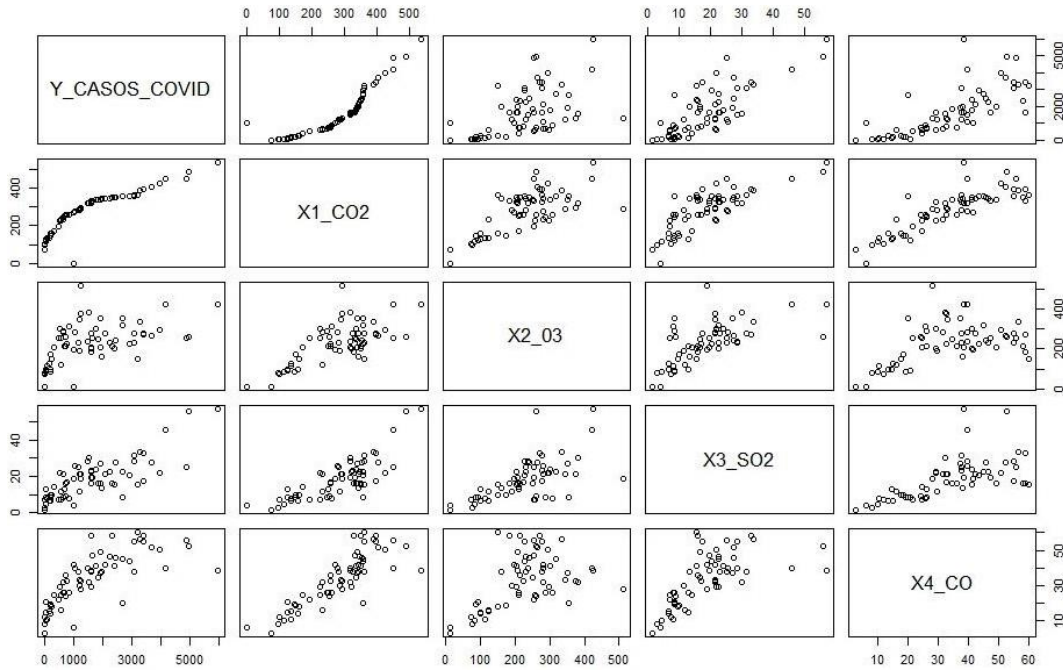
#### 4.9 Regresión Lineal Multiple En R - Studio

Luego de analizar nuestro caso con regresión lineal simple, procedamos en esta sección analizar nuestro modelo con regresión lineal múltiple para de esta manera mostrar la aplicación a los diferentes casos de prueba y del enfoque de solución.

##### 4.9.1 Relación entre variables

Al analizar los siguientes datos es importante manifestar que cuando el aumento de una de las variables viene acompañado del aumento de la otra, se trata de una correlación positiva o directa. En nuestro caso el aumento considerable de los gases contaminantes está correlacionados positivamente. Si el aumento de una variable viene acompañado de una disminución en la otra, la correlación es negativa o inversa.

El hecho de que dos variables estén correlacionadas no quiere decir que una sea la causante de las otras dos variables pueden estar independientemente relacionadas con una tercera (quizás no identificada), es por ello que los contaminantes seleccionados son los que se presentan en gran medida en nuestra atmosfera, se debe considera que al querer obtener resultados se debe considerar varios aspectos que implican el motivo del avance de la enfermedad.



**Figura 35-4.** Matriz de correlaciones

Realizado por: Pérez (2022)

En la figura 4-27 podemos notar que en referencia a la variable Y correspondiente a los casos de Covid-19, todas las variables presentaron una correlación alta, a excepción de la variable 2, la cual arrojó una correlación moderada con respecto a la variable X2\_O<sub>3</sub>.

Para descartar los valores dispersos aplicamos una prueba de Pearson la cual no permitirá tener una idea más clara de la correlación entre las variables independientes respecto a la variable dependiente.

#### 4.9.2 Correlación de Pearson

El análisis de Casos Covid y el contaminante X1\_Co2 es de 0.8988808 o el 89% de correlación, de la misma manera Casos Covid y el contaminante X2\_O<sub>3</sub> mantiene una correlación de 0.5163088 o el 51%, al verificar los Casos Covid y la variable X3\_SO<sub>2</sub> 0.7853471 o el 78%, por último la correlación entre Casos Covid y CO es de 0.754669 o 75 %, podemos manifestar que lo visto en la matriz refleja en la correlación de Pearson para cada contaminante, evidenciado que O<sub>3</sub> es una variable que aporta muy poca información al comportamiento de la variable dependiente Y. Por lo tanto a media que aumenta CO<sub>2</sub>, SO<sub>2</sub>, CO aumenta los casos Covid.

```
Pearson's product-moment correlation
data: Y_CASOS_COVID and X1_CO2
t = 16.152, df = 62, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8382957 0.9375380
sample estimates:
      cor
0.8988808
```

**Figura 36-4.** Correlación de Pearson CASOS COVID y X1\_CO2

Realizado por: Pérez (2022)

```
Pearson's product-moment correlation
data: Y_CASOS_COVID and X2_O3
t = 4.7471, df = 62, p-value = 1.259e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3098198 0.6762880
sample estimates:
      cor
0.5163088
```

**Figura 37-4.** Correlación de Pearson CASOS COVID y X2\_O3

Realizado por: Pérez (2022)

```
Pearson's product-moment correlation
data: Y_CASOS_COVID and X3_SO2
t = 9.9891, df = 62, p-value = 1.556e-14
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6686101 0.8643059
sample estimates:
      cor
0.7853471
```

**Figura 38-4.** Correlación de Pearson CASOS COVID y X3\_SO2

Realizado por: Pérez (2022)

```
Pearson's product-moment correlation
data: Y_CASOS_COVID and X4_CO
t = 9.0569, df = 62, p-value = 5.938e-13
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6247546 0.8439261
sample estimates:
      cor
0.754669
```

**Figura 39-4.** Correlación de Pearson CASOS COVID y X4\_CO

Realizado por: Pérez (2022)



### 4.9.3 Creación de modelo

En este punto, se presentan aplicaciones de los resultados obtenidos en el R, dando respuesta a los tres objetivos específicos planteados en un inicio de cada uno de los análisis realizados, para ello es importante analizar por medio de regresión lineal múltiple, la misma que nos permite analizar los datos y presentar el mejor modelo.

Call:					
lm(formula = Y_CASOS_COVID ~ X1_CO2 + X2_O3 + X3_SO2 - X4_CO, data = DAT)					
Residuals:					
Min	1Q	Median	3Q	Max	
-824.26	-366.61	-66.54	281.95	2226.20	
Coefficients:					
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-1277.0928	205.4016	-6.218	5.31e-08	***
X1_CO2	11.1689	1.1424	9.777	4.99e-14	***
X2_O3	-3.4576	0.9837	-3.515	0.000843	***
X3_SO2	30.2787	10.2183	2.963	0.004360	**
---					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 543.8 on 60 degrees of freedom					
Multiple R-squared: 0.8526, Adjusted R-squared: 0.8453					
F-statistic: 115.7 on 3 and 60 DF, p-value: < 2.2e-16					

**Figura 40-4.** Valores de regresión lineal múltiple

Realizado por: Pérez (2022)

De acuerdo con la tabla obtenida del análisis de datos tenemos la creación del modelo, considerando las variables independientes y dependiente este análisis de correlación de variables se tiene que, el intercepto es de  $-1277.0928$ , el  $\beta_1 = 11.1689$  el valor es positivo por lo tanto los casos aumenta cuando mayor es la concentración del contaminante,  $\beta_2 = -3.4576$ ,  $\beta_3 = 30.2787$ , el modelo por regresión lineal múltiple depurado se tiene como contaminantes principales a CO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub> estos valores tienen mayor significancia dentro del modelo, los contaminantes en el modelo están con un P-VALOR menor a 0,05 por lo tanto se incluye en nuestro modelo, nuestro  $R^2 = 0.8526$ , sería el 85 % es decir que nuestros valores se encuentran muy cerca de nuestra línea de regresión lineal, nuestro  $R^2$  ajustado = 0.8453 nos ayuda en el análisis considerando el número de valores que tenemos en el análisis, considerando estos valores de  $R^2$  tenemos

una idea de que tan bien ajustado está el modelo lineal, esta estimación nos da fuerza al modelo.

Nuestro estadístico ANOVA (F-STATISTIC) en relación con nuestro P-VALOR es muy por debajo del 0.05 por lo tanto nuestro modelo si explica el comportamiento de los contaminantes en relación con los casos COVID.

A continuación, se muestra la ecuación obtenida.

$$Y = -1277.0928 + 11.1689X_1 - 3.4576X_2 + 30.2787X_3$$

## CONCLUSIONES

- Se construyó la base de datos de los contaminantes del aire PM<sub>2.5</sub>, PM<sub>10</sub>, O<sub>3</sub>, SO<sub>2</sub>, CO y los casos nuevos de Covid - 19 para la ciudad de Quito-Ecuador en el periodo comprendido del 13 de marzo 2020 al 13 de marzo 2021, la misma que nos sirvió para el análisis de los diferentes modelos que se aplicaron para buscar una respuesta al comportamiento de virus en diferentes situaciones.
- Se aplicaron análisis de correlación, en donde se obtuvo que la variable X<sub>2</sub>\_O<sub>3</sub> es la que presenta la menor correlación con respecto a la variable dependiente Y, es decir con los casos de Covid - 19 registrados en la ciudad de Quito. Las demás variables presentaron correlaciones mayores a 0.7, siendo consideradas en un rango de entre moderadas y fuertes.
- Se analizaron los datos en dos softwares; SPSS, Matlab y R, en donde se desarrollaron dos modelos, de los cuales ambos arrojaron valores considerados altos de los coeficientes de determinación, sin embargo, el segundo fue el que se obtuvo un valor más alto en el análisis de ANOVA, es decir una mayor exactitud.
- Para terminar, se realiza simulaciones comparativas entre los casos COVID y los CONTAMINANTES para poder verificar si el estudio realizado se puede utilizar a nivel de todo el país y se concluye que el modelo puede utilizar datos aplicando regresión lineal múltiple la misma que tiene un nivel de correlación más alto entre las variables.

## RECOMENDACIONES

- Se recomienda a los futuros investigadores desarrollar modelos matemáticos considerando un mayor número de datos en un periodo más largo de tiempo, para obtener resultados con menor error.
- Se recomienda al Ministerio de Salud Pública del Ecuador y demás organizaciones de salud considerar los contaminantes atmosféricos en el cálculo y simulación de la propagación del Covid 19 para su estudio y publicación en sitios oficiales para su futuro uso en investigaciones científicas.
- Se recomienda a los futuros investigadores experimentar con más metodologías y softwares que permitan obtener datos más exactos y concretos, para interpretar de manera correcta el comportamiento de las variables en cuestión.
- Cabe destacar que la matemática y la ayuda de sistemas computacionales hacen que los objetivos deseados en diferentes campos de la ciencia sean los óptimos, el presente trabajo es de costo mínimo en inversión económica, considerando que los costos de control de una enfermedad y recurso humanos es muy alto, es por ello que es importante invertir en investigación la misma que permitirá alcanzar resultados más bondadosos.

## GLOSARIO

**Análisis de Correlación y Relación.** - El análisis de correlación consiste en un procedimiento estadístico para determinar si dos variables están relacionadas o no. El resultado del análisis es un coeficiente de correlación que puede tomar valores entre -1 y +1. El signo indica el tipo de correlación entre las dos variables.

**Medio Ambiente.** - El medioambiente es el espacio en el que se desarrolla la vida de los distintos organismos favoreciendo su interacción. En él se encuentran tanto seres vivos como elementos sin vida y otros creados por la mano del hombre.

**Modelo Matemático.** - En ciencias aplicadas y en tecnología, un modelo matemático es uno de los tipos de modelos científicos que emplea algún tipo de formulismo matemático para expresar relaciones, proposiciones sustantivas de hechos, variables, parámetros, entidades y relaciones entre variables de las operaciones, para estudiar comportamientos de sistemas complejos ante situaciones difíciles de observar en la realidad. El término modelización matemática es utilizado también en diseño gráfico cuando se habla de modelos geométricos de los objetos en dos (2D) o tres dimensiones (3D).

**Información.** - Información es el nombre por el que se conoce un conjunto organizado de datos procesados que constituyen un mensaje que cambia el estado de conocimiento del sujeto o sistema que recibe dicho mensaje. Existen diversos enfoques para el estudio de la información

**Enfermedad Viral.** - Los virus causan enfermedades infecciosas comunes como el resfrío común, la gripe y las verrugas. También causan enfermedades graves como el VIH y sida, el ébola y la COVID-19. Los virus son como secuestradores. Invaden las células vivas y normales y las usan para multiplicarse y producir otros virus como ellos.

**Continuidad en el Tiempo.** - Continuidad es un término que se refiere al vínculo que mantienen aquellas cosas que están, de alguna forma, en continuo. Hace un tiempo, el concepto también se empleaba como sinónimo de continuación, aunque hoy este uso es algo arcaico.

**Enfermedades Infecciosas.** - Las enfermedades infecciosas son trastornos causados por organismos, como bacterias, virus, hongos o parásitos. Muchos organismos viven dentro y fuera de nuestros cuerpos. Normalmente son inofensivos o incluso útiles. Pero bajo ciertas condiciones, algunos organismos pueden causar enfermedades

**Prevalencia.** - En el campo de la medicina, una medida del número total de personas en un grupo específico que tienen (o tuvieron) cierta enfermedad, afección o factor de riesgo (como el tabaquismo o la obesidad) en un momento específico o durante un período determinado.

**Incidencia.** - Es la cantidad de casos nuevos de una enfermedad, un síntoma, muerte o lesión que se presenta durante un período de tiempo específico, como un año.

**Variabilidad.** - Variabilidad es la propiedad de aquello que es variable. Este adjetivo, que procede del vocablo latino variabilis, refiere a lo que varía, cambia o se modifica.

**Transmisible.** - Enfermedad infecciosa causada por un agente contagioso que se transmite entre personas por contacto directo con una persona infectada o por un medio indirecto, como la exposición a un vector, animal, fómite, producto o ambiente, o el intercambio de un fluido que está contaminado por el agente contagioso.

**Comorbilidades.** - La "comorbilidad", también conocida como "morbilidad asociada", es un término utilizado para describir dos o más trastornos o enfermedades que ocurren en la misma persona. Pueden ocurrir al mismo tiempo o uno después del otro. La comorbilidad también implica que hay una interacción entre las dos enfermedades que puede empeorar la evolución de ambas.

**Significancia.** - En estadística, un resultado o efecto es estadísticamente significativo cuando es improbable que haya sido debido al azar. Una «diferencia estadísticamente significativa» solamente significa que hay evidencias estadísticas de que hay una diferencia; no significa que la diferencia sea grande, importante o radicalmente diferente.

El nivel de significación de una prueba estadística es un concepto estadístico asociado a la verificación de una hipótesis. En pocas palabras, se define como la probabilidad de tomar la decisión de rechazar la hipótesis nula cuando ésta es verdadera (decisión

conocida como error de tipo I, o *falso positivo*). La decisión se toma a menudo utilizando el valor  $p$ : si el valor  $p$  es inferior al nivel de significación, entonces la hipótesis nula es rechazada. Cuanto menor sea el valor  $p$ , más significativo será el resultado.

En otros términos, el nivel de significación de un contraste de hipótesis es una probabilidad  $p$  tal que la probabilidad de tomar la decisión de rechazar la hipótesis nula —cuando esta es verdadera— no es mayor que  $p$ .

**Contaminantes Atmosféricos.** - Los contaminantes primarios son aquellas sustancias que son vertidas directamente a la atmósfera por fuentes emisoras tales como chimeneas, procesos industriales, quema de combustibles fósiles, etc. Estas sustancias representan más del 90% de la contaminación atmosférica.

**ANOVA.** - Análisis de la Varianza ( ANOVA ) es una fórmula estadística que se utiliza para comparar las varianzas entre las medias (o el promedio) de diferentes grupos. Una variedad de contextos lo utilizan para determinar si existe alguna diferencia entre las medias de los diferentes grupos.

**Relación entre variables.** - Si existe relación entre variables significa que existe covariación, es decir, que el modo en que varían es más o menos conjunto. En otras palabras: observamos que cuando se da variación en una variable y en un sentido concreto, también se da variación en la otra u otras variables y también en un sentido concreto.

**Correlación de Pearson.** - El coeficiente de correlación de Pearson es una prueba que mide la relación estadística entre dos variables continuas. Si la asociación entre los elementos no es lineal, entonces el coeficiente no se encuentra representado adecuadamente. El coeficiente de correlación puede tomar un rango de valores de +1 a -1. Un valor de 0 indica que no hay asociación entre las dos variables. Un valor mayor que 0 indica una asociación positiva. Es decir, a medida que aumenta el valor de una variable, también lo hace el valor de la otra. Un valor menor que 0 indica una asociación negativa; es decir, a medida que aumenta el valor de una variable, el valor de la otra disminuye.

## **BIBLIOGRAFÍA**

- Amat Rodrigo, J. (Junio de 2016). *rpubs*. [https://rpubs.com/Joaquin\\_AR/223351](https://rpubs.com/Joaquin_AR/223351)
- Arroyo, I., Bravo M, L. C., Llinás, H., & Muñoz, F. L. (2014). Distribuciones Poisson y Gamma: Una Discreta y Continua Relación. *Prospect*, 99-107.
- Bernal, C. (2006). *Metodología de la investigación. Para administración, economía, humanidades y ciencias sociales*. México: Pearson Educación.
- Bueno, M. (2013). *Modelos Lineales Generalizados para evaluar el efecto de las hormonas reguladoras auxinas y citoquininas en el crecimiento de Gypsophilapaniculata L., 2010*. Trujillo - Perú.
- García, H. A., Solarte, C., & Imuez, M. (2009). *BioEstadística*. Argentina: Universitaria
- Gómez, S. (2012). *Metodología de la Investigación*. México: Red Tercer Milenio CS.
- Gómez, S., & Palacios, D. (2013). *Modelación logística Multinomial para clasificar los hogares de el Salvador*. Salvador : Universidad de El Salvador.
- Gonzalez, A. (2015). *Selección de variables: Una revisión de métodos existentes*. Coruña: Universidad de Coruña.
- Grajales, T. (2000). Tipos de investigación. *Recuperado el 06 de 04 de 2012, de Página web de documentos académicos de Tevni Grajales G.: <http://tgrajales.net/investipos.pdf>, 2.*
- Landeau, R. (2017). *Elaboración de trabajos de investigación*. Caracas, Venezuela: Editorial Alfa.
- Lopez, A. (2004). *Modelos Lineales Generalizados*. Valencia: Universidad de Valencia
- Martínez, M. (2021). Factores determinantes de la rentabilidad empresarial del sector transporte en Colombia. *UNIVERSIDAD EAFIT*, 58.



- Martínez, R. (2009). El coeficiente de concreciones de los rangos de caracterización de Spearman. *Revista Habanera de Ciencias Médicas*, 8(2).
- Mejía, E. (2021). Mejía (2021), cuyo tema es: “Planeación financiera como herramienta de gestión para mejorar la rentabilidad de la Empresa Textil Incofecsa ubicada en la ciudad de Riobamba, periodo 2020-2023”. *Escuela Superior Politécnica de Chimborazo*, 95.
- Ministerio de Deportes del Ecuador. (2016). Lineamientos Para La Presentación De La Planificación Operativa Anual 2017, Ejecución, Seguimiento Y Evaluación Del Plan Operativo Anual 2017 De Las Organizaciones Deportivas Quito,. *Coordinación General De Planificación* , 10.
- Ministerio de Salud Pública. (2022). *Informes de Situación (SITREP) e Infografías COVID-19*. Quito.
- Monjarás, Bazán, Pacheco, Rivera, Zamarripa, & Cuevas. (2019). Diseños de Investigación. *Publicación semestral, Educación y Salud Boletín Científico Instituto de Ciencias de la Salud Universidad Autónoma del Estado de Hidalgo, Vol. 8, No. 15* , 119-122.
- Morales, J. (2018). Modelos Estadísticos. Grado Biotecnología. *Amazonaws*.
- Pita, S., & Pértegas, S. (2002). *Investigación cuantitativa y cualitativa*. Coruña: Cad Aten Primaria .
- Rodríguez, H. (2020). Análisis de la rentabilidad aplicando el modelo DUPONT en empresas de transporte de carga pesada en la provincia del Carchi. . *Revsita Sathiri:Sembrador*, 15(2), 9 - 21. <https://doi.org/10.32645/13906925.976>, 19.
- Roque Cruz , M. J. (2018). *Modelos de Regresión Logística Multinomial de la Calidad de Fibra de Alpaca Huacaya en función de sus características:Sexo y Edad-Corani,Carabaya,Puno- 2017*. Perú : UNIVERSIDAD NACIONAL DEL ALTIPLANO Universidad Nacional del Altiplano .

Salas, Á. (2021). Retrospectiva financiera de la Cooperativa de Ahorro y Crédito 4 de Octubre, periodo 2017 – 2019, mediante la aplicación de la Metodología DuPont. *Escuela Superior Politécnica de Chimborazo*, 64.

Secretaría de Ambiente del Distrito Metropolitano de Quito. (2022). *Red de Monitoreo Atmosférico*. Quito.

Tejeira, J. (2020). Analysis of a simple mathematical model for describing the COVID-19 spreading. *Revista Boliviana de Física*, 37(37), 41-43.

UCLA. (28 de Febrero de 2020). *Statistical Consulting*.  
<https://stats.idre.ucla.edu/r/dae/poisson-regression/>

Zurita Herrera, G. (2010). *Probabilidad y Estadística Fundamentos y Aplicaciones*. Guayaquil : ESPOL.

Zylberberg , A. D. (2005). *Probabilidad y Estadística*. Argentina : Nueva Librería.

## ANEXOS

### ANEXO A: CÓDIGO MATLAB ANÁLISIS REGRESIÓN LINEAL SIMPLE

```
close all, clear all, clc
tabla2=xlsread('datos')
Y=tabla2(:,1); %Covid-19
X1=tabla2(:,2); %CO2
X2=tabla2(:,3);
X3=tabla2(:,4);
X4=tabla2(:,5);
format long
b1=X1\Y
yCalc1=b1*X1;
scatter(X1,Y)
hold on
plot(X1,yCalc1)
xlabel('CO2')
ylabel('Casos de Covid-19')
title('Correlación lineal')
grid on
Rsqr1=1-sum((Y-yCalc1).^2)/sum((Y-mean(Y)).^2)
% Ajuste
m=[ones(length(X1),1),X1];
b=m\Y;
yCalc2=m*b;
plot(m,yCalc2,'--')
legend('Datos','Real','Ajuste').
Rsqr2=1-sum((Y-yCalc2).^2)/sum((Y-mean(Y)).^2)
```

## ANEXO B: CÓDIGO MATLAB REGRESIÓN LINEAL MÚLTIPLE

```
```{r}REGRESION LINEAL
```

```
library(readxl)
```

```
install.packages("tidyverse")
```

```
library(tidyverse)
```

```
library(car)
```

```
library(boot)
```

```
library(QuantPsyc)
```

```
library(ggplot2)
```

```
sole=read_excel("TESIS R/BASE_DE_DATOS_FINAL.xlsx")
```

```
sole
```

```
modlo1=lm(Y_CASOS_COVID~X1_CO2+X2_03+X3_SO2-X4_CO, data = sole)
```

```
summary(modlo1)
```

```
sole <- read_excel("TESIS R/BASE_DE_DATOS_FINAL.xlsx")
```

```
View(sole)
```

```
attach(sole)
```

```
names(sole)
```

```
class(Y_CASOS_COVID)
```

```
class(X1_CO2)
```

```
class(X2_03)
```

```
class(X3_SO2)
```

```
class(X4_CO)
```

```
modelo1 = lm (Y_CASOS_COVID~X1_CO2 , data = sole, na.action = na.exclude)
```

```
summary(modelo1)
```

```
modelo1$coefficients
```

```

sqrt(0.808)
grafica1 = ggplot(sole,aes(X1_CO2,Y_CASOS_COVID))
grafica1+geom_point()
grafica1+geom_point()+geom_smooth(method="lm",colour="blue")
modelo2 = lm (Y_CASOS_COVID~X2_03 , data = sole, na.action = na.exclude)
summary(modelo2)
modelo1$coefficients
sqrt(0.2666)
grafica1 = ggplot(sole,aes(X2_03,Y_CASOS_COVID))
grafica1+geom_point()
grafica1+geom_point()+geom_smooth(method="lm",colour="blue")
modelo3 = lm (Y_CASOS_COVID~X3_SO2 , data = sole, na.action = na.exclude)
summary(modelo3)
modelo1$coefficients
sqrt(0.6168)
grafica1 = ggplot(sole,aes(X3_SO2,Y_CASOS_COVID))
grafica1+geom_point()
grafica1+geom_point()+geom_smooth(method="lm",colour="blue")
modelo4 = lm (Y_CASOS_COVID~X4_CO , data = sole, na.action = na.exclude)
summary(modelo4)
modelo1$coefficients
sqrt(0.5695)
grafica1 = ggplot(sole,aes(X4_CO,Y_CASOS_COVID))
grafica1+geom_point()
grafica1+geom_point()+geom_smooth(method="lm",colour="blue")

```

**Anexo 1. METODOLOGÍA stepAIC (MEJOR MODELO)**

```

library(MASS)
stepAIC(modlo1,method=c("both"))
modlo2=lm(Y_CASOS_COVID~X1_CO2+X2_03+X3_SO2,data = sole)
summary(modlo2)

```

## **ANEXO C: SHAPIRO TEST**

```
library(car)
```

```
shapiro.test(modlo2$residuals)
```

```
durbinWatsonTest(modlo2)
```

```
ncvTest(modlo2)
```

```
vif(modlo2)
```

## **ANEXO D: PRUEBA DE NORMALIDAD**

**kolmogorov smirnov (n>=50)**

Kolmogorov-Smirnov

library(nortest)

lillie.test(modlo2\$residuals)

qqnorm(modlo2\$residuals)

qqline(modlo2\$residuals)

boxplot(modlo2\$residuals)

## **ANEXO E: IDEPENDENCIA DE LOS ERRORES**

```
library("readxl")
```

```
library("car")
```

```
durbinWatsonTest(modlo2)
```

```
ncvTest(modlo2)
```

```
vif(modlo2)
```

### **No hay multicolinealidad**

```
dataset <- read_excel("TESIS R/BASE_DE_DATOS_FINAL.xlsx")
```

```
```{r}
```

```
sapply(dataset, class)
```

```
```
```

```
regresion1 <- lm(Y_CASOS_COVID~., dataset)
```

```
summary(regresion1)
```

```
regresion2 <- lm(Y_CASOS_COVID~X1_CO2+X2_03, dataset)
```

```
summary(regresion2)
```

```
regresion2 <- lm(Y_CASOS_COVID~X2_03+X3_SO2, dataset)
```

```
summary(regresion2)
```

```
regresion2 <- lm(Y_CASOS_COVID~X3_SO2+X4_CO, dataset)
```

```
summary(regresion2)
```

```
library("readxl")
```

```
data.frame <- read_excel("TESIS R/BASE_DE_DATOS_FINAL.xlsx")
```

```
data_pred <- data.frame(X1=c(431.00,453.00,245.00,354.00,45.00,190.00,230.00,  
269.00,262.00,97.00,287.00,254.00,238.00,172.00,112.00,265.00,357.00,278.00,296.00,  
,106.00,283.00,371.00,240.00,328.00,139.00,232.00,285.00,329.00,336.00,154.00,285.  
00,288.00,396.00,137.00,0.00,319.00,296.00,332.00,347.00,159.00,458.00,547.00,386.  
00,394.00,192.00,362.00,413.00,350.00,342.00,125.00,364.00,349.00,343.00,357.00,13  
5.00,342.00,352.00,255.00,338.00,333.00,367.00,343.00,398.00,155.00
```

```
),X2=c(314.00,271.00,299.00,360.00,0.00,255.00,259.00,310.00,269.00,78.00,192.00,2  
12.00,119.00,210.00,87.00,229.00,285.00,194.00,230.00,75.00,209.00,283.00,305.00,2  
40.00,101.00,290.00,284.00,290.00,366.00,178.00,352.00,388.00,349.00,123.00,13.00,  
397.00,525.00,295.00,333.00,146.00,419.00,429.00,261.00,284.00,104.00,239.00,274.0  
0,247.00,214.00,98.00,208.00,238.00,261.00,236.00,129.00,266.00,170.00,227.00,224.
```



```
00,170.00,164.00,233.00,303.00,93.00),X3=c(19.00,22.00,8.00,8.00,2.00,6.00,7.00,7.00
,9.00,2.00,11.00,11.00,12.00,13.00,4.00,17.00,21.00,19.00,25.00,7.00,21.00,33.00,21.0
0,29.00,14.00,21.00,25.00,20.00,26.00,10.00,21.00,21.00,34.00,8.00,4.00,29.00,14.00,2
3.00,23.00,9.00,42.00,56.00,61.00,34.00,7.00,29.00,28.00,27.00,21.00,9.00,18.00,24.00
,16.00,21.00,6.00,20.00,14.00,16.00,16.00,16.00,16.00,16.00,20.00,9.00),X4=c(52.00,6
1.00,28.00,21.00,3.00,22.00,25.00,33.00,31.00,10.00,35.00,36.00,18.00,36.00,13.00,35.
00,49.00,41.00,47.00,15.00,18.00,26.00,20.00,45.00,15.00,31.00,36.00,37.00,37.00,17.
00,32.00,30.00,56.00,10.00,4.00,32.00,25.00,41.00,43.00,16.00,41.00,37.00,53.00,58.0
0,13.00,51.00,57.00,51.00,36.00,19.00,37.00,43.00,33.00,44.00,14.00,49.00,42.00,33.0
0,45.00,59.00,64.00,57.00,51.00,19.00))
```

```
predict(regresion2,data_pred)
```

```
par(mfrow=c(1,5))
```

```
for (i in 2:5) {boxplot(dataset[,i], main=names(dataset)[i])}
```

```
hist(dataset$X1_CO2)
```

```
hist(log(dataset$X2_03))
```

```
hist(log(dataset$X3_SO2))
```

## ANEXO F: ANÁLISIS DE CORRELACIÓN ENTRE DOS VARIABLES

```
```{r}
plot(dataset$X1_CO2, dataset$Y_CASOS_COVID)
```
```{r}
plot(dataset$X2_03, dataset$Y_CASOS_COVID)
```
```{r}
plot(dataset$X3_SO2, dataset$Y_CASOS_COVID)
```
```{r}
plot(dataset$X4_CO, dataset$Y_CASOS_COVID)
```
```{r}
plot(log(dataset$X1_CO2), dataset$Y_CASOS_COVID)
```
```{r}
plot(log(dataset$X2_03), dataset$Y_CASOS_COVID)
```
```{r}
plot(log(dataset$X3_SO2), dataset$Y_CASOS_COVID)
```
```{r}
plot(log(dataset$X4_CO), dataset$Y_CASOS_COVID)
```
library(ggplot2)
ggplot(dataset, aes(log(X2_03), Y))+geom_point()+stat_smooth(method =
'lm', col="blue")+labs(title = "Regresion lineal simple", x = "X:variable independiente",
y = "Y:variable dependiente", caption = 'Fuente: elaboracion propia') + theme_gray()
```{r}
library(ggExtra)
```

## **ANEXO G: GRÁFICO CENTRAL CON GGLOT CORRELACIONES**

```
plt_1<-ggplot(dataset, aes(log(X3_SO2),Y_CASOS_COVID)) + geom_point(col =  
'#00a0dc') + stat_smooth(method = 'lm', col = '#ec4339') + labs(title = "Regresion lineal  
simple", x="X:variable independiente", y="Y:variable dependiente", caption = 'Fuente:  
Elaboracion propia')+theme_gray()
```

```
summary(plt_1)
```

## **ANEXO H: ADICIÓN DE DISTRIBUCIONES**

```
ggMarginal(plt_1, type="dens")
```

```
library(GGally)
```

```
ggpairs(data)
```

### **POISSON**

```
install.packages("ISwR")
```

```
library(ISwR)
```

```
PoissonModel1 <- glm(Y_CASOS_COVID~X2_03+X3_SO2,data = sole)
```

```
summary(PoissonModel1)
```

## **ANEXO I: REGRESIÓN LINEAL MULTIPLE**

```
install.packages("tidyverse")
```

```
install.packages("lmtest")
```

```
library(readxl)
```

```
library(tidyverse)
```

```
library(car)
```

```
library(boot)
```

```
library(QuantPsyc)
```

```
library(ggplot2)
```

```
library(lmtest)
```

## **ANEXO J: CREACIÓN DE MODELO**

```
DAT=read_excel("TESIS R/BASE_DE_DATOS_FINAL.xlsx")
```

```
SUMMARY=DAT
```

```
attach(DAT) #accede a las variables
```

```
pairs(DAT) #selecciona las variables
```

```
#options(scipen = 999)#quita notacion cientifica
```

```
#relacion entre variables
```

```
cor(Y_CASOS_COVID,X1_CO2)#matriz de correlacion
```

```
cor(Y_CASOS_COVID,X2_03)#matriz de correlacion
```

```
cor(Y_CASOS_COVID,X3_SO2)#matriz de correlacion
```

```
cor(Y_CASOS_COVID,X4_CO)#matriz de correlacion
```

```
cov(Y_CASOS_COVID,X1_CO2)#covarianza entre casos covid
```

```
cov(Y_CASOS_COVID,X2_03)#covarianza entre casos covid
```

```
cov(Y_CASOS_COVID,X3_SO2)#covarianza entre casos covid
```

```
cov(Y_CASOS_COVID,X4_CO)#covarianza entre casos covid
```

```
cor.test(Y_CASOS_COVID,X1_CO2)#test de correlacion de pearson
```

```
cor.test(Y_CASOS_COVID,X2_03)#test de correlacion de pearson
```

```
cor.test(Y_CASOS_COVID,X3_SO2)#test de correlacion de pearson
```

```
cor.test(Y_CASOS_COVID,X4_CO)#test de correlacion de pearson
```

## ANEXO K: SELECCIÓN DE MODELO

```
RM <- lm(Y_CASOS_COVID~X1_CO2+X2_03+X3_SO2-X4_CO, data = DAT)
```

```
summary(RM)
```

```
plot(RM) #RESIDUAL DE VARIANZAS GRAFICOS DE NORMALIDAD
```

```
residuos<-residuals(RM)
```

```
hist(residuos)
```

```
plot(RM,)
```

```
require(lmtest)
```

```
dwtest(RM)
```

```
vif (RM)
```

TABLA 1-1. Descripción del Presupuesto

RUBROS	CANTIDAD	UNIDAD DE MEDIDA	COSTO UNITARIO	COSTO TOTAL
Papel	3	Resmas	3	9
Fotocopias	200	Unidades	0.05	10
Internet	20	Hora	1	20
Impresión del anteproyecto	1	unidades	20	20
Impresión Predefensa	3	Unidades	30	90
Impresión defensa	3	Unidades	40	120
SPSS	1	Unidades	80	80
MATLAB	1	Unidades	180	180
R- studio	1	Unidades	150	150
TOTAL				679

**Elaborado por:** Pérez (2022)

Nota. Se tomaron en cuenta todos los rubros que se necesitaron para obtener el modelo matemático deseado.



epoch

Dirección de Bibliotecas y  
Recursos del Aprendizaje

UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y  
DOCUMENTAL

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 22 / 11 / 2022

<b>INFORMACIÓN DEL AUTOR/A (S)</b>
<b>Nombres – Apellidos:</b> <i>Carlos Gabriel Pérez Castillo</i>
<b>INFORMACIÓN INSTITUCIONAL</b>
<i>Instituto de Posgrado y Educación Continua</i>
<b>Título a optar:</b> <i>Magíster en Matemática mención Modelación y Docencia</i>
<b>f. Analista de Biblioteca responsable:</b> Lic. Luis Caminos Vargas Mgs.



Firmado electrónicamente por:  
**LUIS ALBERTO  
CAMINOS  
VARGAS**



0169-DBRA-UTP-IPEC-2022



• TRADUCCION

---



• **JAZMINA IVONNE MENA MAYORGA** <jazmina.mena@epoch.edu.ec>  
Para: calin86gpec@yahoo.es  
CC: Centro de Idiomas

Con un cordial saludo, adjunto documento traducido.

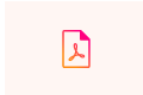
Jazmin

***"Educating the mind without educating the heart, is no education at all" Aristotle***

[📄](#) Descargar todos los archivos adjuntos como archivo comprimido



CALIN\_PER...docx  
13.4kB



CALIN\_PERE...pdf  
59.4kB

