



Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Estado de México

Avenida: Ciencias Aplicadas

Reporte. “Los factores meteorológicos y la concentración de los contaminantes en la atmósfera”

Equipo 3. Integrantes:

Tania Camila Castillo Robles	A01660703
Ana Martínez Barbosa	A01382889
Elisa Sánchez Bahnsen	A01745371
Cristian Gonzaga López	A01745134
Adrián Landaverde Nava	A01745052

Unidad de formación:

MA 1031.308. Análisis estadístico

Semestre 2 - 2021

Maestro:

Raúl Martínez Rosado

Fecha de entrega:

10 de junio de 2021

1.- ACTIVACIÓN:

a.- Contexto.

Responsable: Tania Camila Castillo Robles

La atmósfera es un sistema dinámico cuya fuente principal de energía es el Sol y causa primaria de la mayoría de los procesos que tienen lugar en el planeta (desde la circulación general atmosférica hasta la propia vida). La energía recibida es aproximadamente de $3,7 \times 10^{21}$ calorías/día. Esta energía llega en forma de radiación electromagnética, abarcando todo el espectro que va desde los rayos gamma (pasando por los rayos X, ultravioleta, visible, infrarrojo, microondas) hasta las radiofrecuencias. Una gran parte de esta radiación, que sería muy perjudicial para los seres vivos, es absorbida por la atmósfera, que actúa como escudo protector contra esas radiaciones.

La contaminación atmosférica es la presencia en la atmósfera de sustancias, materias o formas de energía que supongan una molestia grave, un peligro o un daño para el ser humano o el medio ambiente, según recoge el Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente.

Son diversas las causas que explican la contaminación de la atmósfera, tales como las extracciones mineras, uso de pesticidas en las actividades agrícolas, industrialización excesiva, quema de combustibles, la deforestación, etcétera.

Son muchos los autores que han demostrado la estrecha relación existente entre las situaciones meteorológicas y los niveles de contaminación alcanzados en un momento dado. Estas situaciones atmosféricas vienen determinadas por los valores que toman ciertas variables, tales como la presión, la temperatura, la humedad, etc. Por otra parte, la dispersión de estos contaminantes es influenciada por la velocidad y dirección de los vientos, así como la estabilidad que predomina en la atmósfera.

En México, se ha puesto atención al problema de la contaminación atmosférica, debido principalmente a las actividades petroleras que se realizan en el suroeste. Con ayuda del Sistema Automatizado de la Región Sur (SAMARS), red de monitoreo autorizada, gracias a Petróleos Mexicanos (PEMEX) se pueden monitorear los principales contaminantes atmosféricos a excepción del ozono y variables meteorológicas. Esta red cuenta con seis estaciones de monitoreo ubicadas en las periferias de las instalaciones petroleras.

Nuestro objetivo principal es poder identificar aquella relación que tiene las variables meteorológicas, tales como la humedad relativa, la temperatura y la radiación solar y la concentración de contaminantes en la atmósfera, así como poder entender cómo afectan estas variables a la concentración de contaminantes.

2.- ADQUISICIÓN:

a.- Indagación

Responsable: Cristian Gonzaga López

Entre los principales contaminantes que se encuentran en la atmósfera, los más presentes son el dióxido de azufre, el dióxido de nitrógeno y el dióxido de carbono. Debido a la relevancia actual de la búsqueda necesaria de una reducción en contaminación ambiental, diversos estudios se han realizado para encontrar factores que pudiesen estar relacionados en dicho suceso.

Por ejemplo, se ha presentado una relevancia significativa que muestra que factores tales como la presión, la temperatura ambiental, la precipitación y la ubicación de un lugar pueden ayudar a aumentar o disminuir la concentración de contaminantes(Hayas,1977).

Entre los resultados obtenidos de algunos de dichos estudios, se encontró que el viento de un lugar particular juega un rol crítico en su concentración de contaminantes. Sin embargo, cabe resaltar que los flujos de aire se originan dependiendo de la temperatura del lugar y la presión. Por lo que tanto la temperatura como la presión atmosférica están involucrados en dicha correlación (Hayas,1991).

El viento de un lugar permite el movimiento del aire contaminado dando lugar así a la eliminación natural de contaminantes atmosféricos. Sin embargo, la temperatura puede jugar en su favor o en su contra. En el caso de la inversión térmica, hay ciudades en las que las bajas temperaturas impiden dicha eliminación natural debido a que no puede fluir por su densidad (Coppel, 2017). Por otra parte, el verano puede ser perjudicial al facilitar la activación del ozono troposférico en zonas tanto urbanas como rurales.

En relación con el viento, otro factor meteorológico que afecta la concentración de contaminantes es la presión atmosférica. Esta última se define como el peso que tiene la atmósfera sobre nosotros en cierta posición (Santos, 2017) y es relevante puesto que en zonas de alta presión se llevan a cabo los anticiclones. Ahora bien, la razón por la cual los anticiclones se presentan como causas de la concentración de contaminantes es debido a la dirección del viento durante ese fenómeno pues el anticiclón empuja el aire hacia abajo impidiendo la dispersión de los contaminantes hacia lo alto de la atmósfera.

b.- Datos

Responsable: Ana Martínez Barbosa

El monitoreo de la calidad de aire se realiza por el Sistema de Monitoreo Atmosférico y consiste en distintas mediciones de contaminantes en una serie de estaciones ubicadas en el interior de la Zona Metropolitana de la Ciudad de México. Dichas estaciones de Monitoreo son:

Clave	Nombre	Alcaldía o municipio	Entidad
ACO	Acolman	Acolman	Estado de México
AJU	Ajusco	Tlalpan	CDMX
AJM	Ajusco Medio	Tlalpan	CDMX
ATI	Atizapán	Atizapán de Zaragoza	Estado de México
BJU	Benito Juárez	Benito Juárez	CDMX
CAM	Camarones	Azcapotzalco	CDMX
CCA	Centro de Ciencias de la Atmósfera	Coyoacán	CDMX
TEC	Cerro del Tepeyac	Gustavo A. Madero	CDMX
CHO	Chalco	Chalco	Estado de México
COR	CORENA	Xochimilco	CDMX
CUA	Cuajimalpa	Cuajimalpa de Morelos	CDMX
CUT	Cuautitlán	Tepotztlán	Estado de México
DIC	Diconsa	Tlalpan	CDMX
EAJ	Ecoguardas Ajusco	Tlalpan	CDMX
EDL	Ex Convento Desierto de los Leones	Cuajimalpa de Morelos	CDMX
FAC	FES Acatlán	Naucalpan de Juárez	Estado de México
FAR	FES Aragón	Nezahualcóyotl	Estado de México
GAM	Gustavo A. Madero	Gustavo A. Madero	CDMX
HGM	Hospital General de México	Cuauhtémoc	CDMX
INN	Investigaciones Nucleares	Ocoyoacac	Estado de México
IZT	Iztacalco	Iztacalco	CDMX
LPR	La Presa	Tlalnepantla de Baz	Estado de México
LAA	Laboratorio de Análisis Ambiental	Gustavo A. Madero	CDMX

LAA	Laboratorio de Análisis Ambiental	Gustavo A. Madero	CDMX
IBM	Legaria	Miguel Hidalgo	CDMX
LOM	Lomas	Miguel Hidalgo	CDMX
LLA	Los Laureles	Ecatepec de Morelos	Estado de México
MER	Merced	Venustiano Carranza	CDMX
MGH	Miguel Hidalgo	Miguel Hidalgo	CDMX
MPA	Milpa Alta	Milpa Alta	CDMX
MON	Montecillo	Texcoco	Estado de México
MCM	Museo de la Ciudad de México	Cuauhtémoc	CDMX
NEZ	Nezahualcóyotl	Nezahualcóyotl	Estado de México
PED	Pedregal	Álvaro Obregón	CDMX
SAG	San Agustín	Ecatepec de Morelos	Estado de México
SNT	San Nicolás Totolapan	La Magdalena Contreras	CDMX
SFE	Santa Fe	Cuajimalpa de Morelos	CDMX
SAC	Santiago Acahualtepec	Iztapalapa	CDMX
TAH	Tláhuac	Xochimilco	CDMX
TLA	Tlalnepantla	Tlalnepantla de Baz	Estado de México
TLI	Tultitlán	Tultitlán	Estado de México
UIZ	UAM Iztapalapa	Iztapalapa	CDMX
UAX	UAM Xochimilco	Coyoacán	CDMX
VIF	Villa de las Flores	Coacalco de Berriozábal	Estado de México
XAL	Xalostoc	Ecatepec de Morelos	Estado de México

Figura 1. Estaciones de Monitoreo.

Adaptado de Aire CDMX [Fotografía], por Sistema de Monitoreo Atmosférico, 2021 (<http://www.aire.cdmx.gob.mx/default.php?opc=%27ZaBhnmI=&dc=%27ZA==>).

- Distribución

Debido a que el monitoreo que se hace es bastante amplio, el Sistema de Monitoreo Atmosférico se divide en 4 distintas redes que se encargan de objetivos específicos sin embargo en conjunto ayudan a

explicar la tendencia de contaminantes atmosféricos. Entre las tareas particulares de cada red se encuentran:

- **Red automática de Monitoreo Atmosférico (RAMA)**

Entre sus funciones más importantes se encuentra evaluar el estado de la calidad del aire en la Zona Metropolitana de la Ciudad de México, evaluar concentración de contaminantes, prevenir a la población de posibles riesgos a través del índice de calidad de aire, activar alertas asociados a un incremento en la concentración de contaminantes y evaluar las estrategias de gestión de calidad de aire de la zona.

- **Red de Depósito Atmosférico (REDDA)**

Se encarga de estudiar la lluvia ácida en la Zona Metropolitana de la Ciudad de México y determinar su impacto en el suelo y en la población.

- **Red de Meteorología y Radiación Solar (REDMET)**

Su principal tarea y objetivo es informar a la población de riesgos asociados a la radiación solar

- **Red Manual de Monitoreo Atmosférico (REDMA)**

Se encarga de evaluar los niveles de exposición de la población ante partículas suspendidas (PM10 y PM2.5) y plomo y asegurarse de que se encuentren dentro de las normas oficiales.

Para la realización del análisis estadístico correspondiente se recurrió a las bases de datos proporcionadas por las redes anteriormente mencionadas. Así mismo, se utilizaron 3 bases de datos, la primera que contenía datos meteorológicos, la segunda que contenía información respecto a la presión atmosférica y la tercera que mostraba la concentración por contaminantes. Para los fines de este estudio se decidió que los únicos parámetros que se analizarían serían la concentración de dióxido de azufre (ppb), la temperatura (Celsius), la presión atmosférica (mmHg) y el porcentaje de humedad relativa(%).

Por otra parte cabe señalar que la información recopilada fue organizada por año y que se posee información desde 2012 hasta los datos más recientes del 2020.

3.- PRÁCTICA:

a.- Desarrollo.

Responsable: Adrián Landaverde Nava

Para llevar a cabo este proyecto, el método por el cual se optó fue R, ya que es un lenguaje de programación especializado en el análisis estadístico. A diferencia de otros softwares como Excel, que también hubieran funcionado para lograr el mismo objetivo, en R se pudieron manipular los datos en cuanto a limpieza y organización por medio de código, mientras que en otros softwares, esto tal vez se hubiera tenido que realizar de forma manual y más lenta volviéndolo menos eficiente. Se menciona lo anterior porque R fue una plataforma que nos proporcionó una gran variedad de herramientas útiles para desechar aquellos datos que no necesitábamos o alteraban los resultados, para organizar los datos en distintas categorías como por ejemplo, por estación, mes o año, y posteriormente para brindarnos

aquella información relacionada al análisis de datos así como los coeficientes de determinación y correlación para lograr esto se usaron librerías especializadas en el manejo y graficación de datos como: tidyverse, dplyr, gapminder y GGally . Adicionalmente, el software nos permitió generar una aplicación web local sencilla con la cual el usuario puede elegir tanto la estación de estudio como también el periodo de tiempo en el cual se quieran visualizar los datos, esto al descargar las librerías shiny y shinydashboard. Por medio de la aplicación, R presenta tanto las gráficas de regresión lineal y múltiple junto con la información estadística dependiendo de la configuración que el usuario desee ingresar. Lograr esto no solo resulta informativo para los usuarios, sino que también fue una manera eficiente en la cual en conjunto pudimos analizar los resultados bajo distintas circunstancias para que finalmente pudiéramos elegir aquellos resultados con los cuales queríamos trabajar. Fue importante asegurarnos que la elección que se tomara, se realizara de manera consciente, ya que nuestro objetivo era encontrar aquel conjunto de datos que al aplicarle los modelos de regresión lineal y múltiple, se explicara exitosamente la relación entre los factores meteorológicos y la contaminación para que por último se generaran las conclusiones finales.

b.- Cálculos.

Responsable: Adrián Landaverde Nava

A partir de los datos obtenidos registrados por el Sistema de Monitoreo Atmosférico de la Zona Metropolitana de la Ciudad de México se buscaba obtener una regresión lineal. El objetivo de la regresión lineal es verificar si se pueden explicar la relación entre dos variables para así poder obtener la ecuación que lo modele y así hacer predicciones o estimaciones sobre los datos. Para esta función lineal simple trabajaremos con la variable dependiente de la concentración de dióxido de azufre (ppb) y la variable independiente será el tiempo, ya que lo que nos interesa es comprender cómo se comporta la concentración de dióxido de azufre a través de los años (de 2009 a 2020) y verificar si existe alguna tendencia.

Para la regresión lineal solo se utilizó las bases de datos referentes a los 12 años de estudio (2012-2020) del apartado “Contaminantes”. Sin embargo antes de poder realizarla se debió de hacer una limpieza y filtración de la base completa de datos ya que no toda su información nos era útil. Por ejemplo, se limpiaron los datos obtenidos de la plataforma ya que en algunos casos se encontraban celdas vacías o errores en la fecha o datos. Posteriormente a la limpieza se realizó la filtración de los datos según lo que se buscaba, en este caso solo se estaba interesado en el dióxido de azufre por lo que fue el único contaminante que se dejó.

Además, cada archivo de datos de cada año incluía la información de todas las estaciones juntas, por lo que se optó por hacer una tabla por cada una de las nueve estaciones por separado. Este filtro y limpieza se realizó para cada año de estudio. Posteriormente se obtuvo el promedio de concentraciones de dióxido de azufre para cada año y se agrupó en una tabla. Cabe resaltar que se hizo una tabla por cada una de las nueve estaciones. A continuación se muestra un ejemplo de las tablas obtenidas:

	date	partes_por_billon	predict(regresion)
1	2011-01-01	2.527397	2.592397
2	2012-01-01	3.019481	2.444460
3	2015-01-01	1.453704	2.000244
4	2016-01-01	1.369565	1.852307
5	2017-01-01	1.935374	1.703965
6	2018-01-01	1.501577	1.556028
7	2019-01-01	1.471299	1.408091
8	2020-01-01	1.539249	1.260154

Figura 2. Ejemplo de tablas obtenidas para cada una de las estaciones.

Ahora que ya se tenía la información ordenada y en el formato requerido, se realizó la regresión lineal para cada una de las estaciones:

```
#-----
# PROMEDIAR POR FECHA MON
#-----
df_MON_RL<- select(df_MON_RL, date, partes_por_billon)
df_MON_RL %>%
  group_by(date) %>%
  tally()
df_MON_RL %>% group_by(date) %>%
  summarise_all(mean)-> df_MON_RL
#-----
# REGRESIÓN LINEAL MON
#-----
regresion=lm(df_MON_RL$partes_por_billon~df_MON_RL$date, data=df_MON_RL)
summary(regresion)
df_MON_RL<- cbind(df_MON_RL,predict(regresion))
ggplot(df_MON_RL, aes(x=date, y=partes_por_billon)) +
  geom_smooth(method="lm", se=FALSE, color="lightgrey") +
  geom_segment(aes(xend=date, yend=predict(regresion)), col='red', lty='dashed') +
  geom_point() +
  geom_point(aes(y=predict(regresion)), col='red') +
  theme_light() +
  ggtitle ("Estación MON: Concentración del dióxido de azufre.") + # Título del gráfico.
  theme(plot.title = element_text(hjust = 0.5)) + # Centrar título
  labs(x = "Año",
       y = "Concentración SO2, ppb")
```

Figura 3. Ejemplo de código para regresión lineal por estación.

Debido a que se realizó la regresión lineal para cada estación, se obtuvieron como resultado 9 tablas diferentes con los datos correspondientes de coeficiente de determinación y análisis general de la estación. Por lo tanto se hizo una selección de aquella estación que tuviese el p-value más pequeño así como el coeficiente de determinación más cercano a 1. Finalmente la estación que cumplió con dichos requisitos fue la estación de Montecillo, Texcoco, Estado de México. Entre los resultados obtenidos cabe destacar que el coeficiente de determinación fue de 0.6109, por lo que se puede concluir que el modelo explica en un 61% la relación entre los datos. Así mismo, en caso de que se deseara hacer un pronóstico para los años consiguientes, se podría utilizar la siguiente ecuación que también fue obtenida de la regresión lineal.

$$y = -0.0004053x + 8.662$$

Donde y representa la concentración de dióxido de azufre (ppb) y x representa el tiempo (año)

Resumen de Regresión Lineal:

```
Cell:
lm(formula = df_MON_RL$partes_por_billon ~ df_MON_RL$date, data = df_MON_RL)

Residuals:
    Min       1Q   Median       3Q      Max
-0.54654 -0.16944  0.00438  0.24333  0.57502

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.6618671   2.2233248    3.896  0.00802 **
df_MON_RL$date -0.0004053   0.0001321   -3.069  0.02196 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4093 on 6 degrees of freedom
Multiple R-squared:  0.6109,    Adjusted R-squared:  0.5461
F-statistic: 9.421 on 1 and 6 DF,  p-value: 0.02196
```

Figura 4. Resultados Regresión lineal estación Montecillo, Texcoco.

Gráfico de Regresión Lineal:

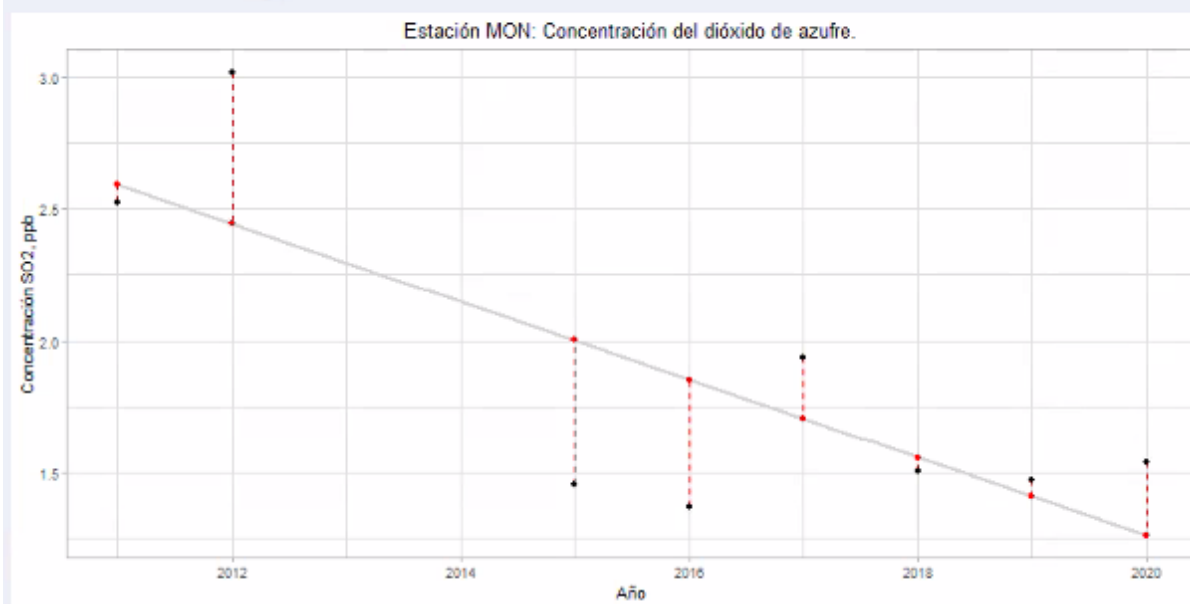


Figura 5. Gráfica regresión lineal estación Montecillo, Texcoco.

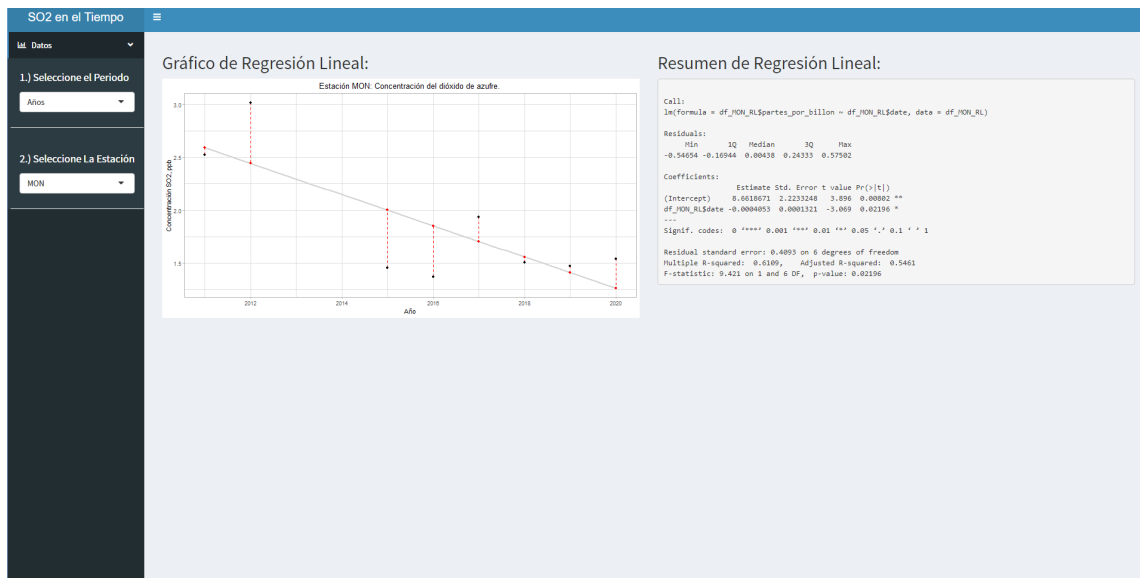


Figura 6. Ejemplo de la aplicación web local para la regresión lineal

En cuanto a la regresión múltiple, de igual manera se obtuvieron los datos desde el 2009 hasta el 2020, haciendo su respectiva limpieza de los datos al obtener aquellos que tuvieran los datos de las partículas y de los datos meteorológicos para esa misma fecha. Además, en esta regresión se comparó la correlación de la cantidad de partículas con el tiempo, la presión, temperatura, humedad, dirección y velocidad del viento. Además, se le dio la libertad al usuario de elegir el contaminante que quisiera comparar.

Ahora que ya se tenía la información ordenada y en el formato requerido, se realizó la regresión múltiple para cada una de la estación y contaminante seleccionado:

	date	partes_por_billon	PA_mmHg	TMP_Celsius	RH_porcentaje	Direccion_viento	Velocidad_viento
1	2012-02-01	5.102273	585.1023	17.80795	44.96591	182.81250	1.731250
2	2012-03-01	4.346723	584.5666	18.46385	38.56237	184.12896	1.779915
3	2012-04-01	3.853309	584.5385	19.30859	33.11270	192.44544	1.899284
4	2012-05-01	6.119241	584.3008	21.01003	32.92412	223.86450	2.149864
5	2012-06-01	2.635607	584.3344	18.75657	55.46589	211.15474	1.844925
6	2012-07-01	3.399361	585.7588	17.35735	64.81629	229.16294	1.875399
7	2012-08-01	2.866667	585.0205	17.70872	63.67179	273.89231	2.450256
8	2013-01-01	14.653979	585.3962	15.64637	42.91003	174.94464	1.661246
9	2013-02-01	7.506234	585.0524	17.76509	33.48628	168.71072	1.555860
10	2013-03-01	8.890110	584.9953	17.15463	34.24490	191.13501	2.016797
11	2013-04-01	5.474910	584.3369	20.79373	32.19176	173.18280	1.814158
12	2013-05-01	4.658333	585.0633	20.10900	44.96667	201.08167	1.826167
13	2013-06-01	5.270769	585.0000	18.93400	55.78769	233.68769	2.034615

Figura 7. Ejemplo de tablas obtenidas para cada una de las estaciones.

```

df_todos_RM$date<- as.Date(df_todos_RM$date, format="%d/%m/%Y")
df_estacion_RM <- df_todos_RM[df_todos_RM$id_station==estacion, ]

df_estacion_RM<- select(df_estacion_RM, date,
                        partes_por_billion,
                        PA_mmHg,
                        TMP_Celsius,
                        RH_porcentaje,
                        Direccion_viento,
                        Velocidad_viento)

df_estacion_RM %>%
  group_by(date) %>%
  tally()
df_estacion_RM%>%group_by(date)%>%
  summarise_all(mean)-> df_estacion_RM

df_estacion_RM$date<- as.numeric(df_estacion_RM$date)

corrplot(cor(df_estacion_RM), type="lower", method="circle")

ggpairs(df_estacion_RM, lower = list(continuous = "smooth"),
        diag = list(continuous = "barDiag"), axisLabels = "none")

print(summary(regresionMultiple))

```

Figura 8. Código para regresión múltiple.

Por lo tanto, una vez obteniendo este código, se procedió a hacer la regresión múltiple con la estación anterior, es decir con la estación MON (Montecillo, Texcoco, Estado de México).

Resumen de Regresión Múltiple:

```

Call:
lm(formula = df_estacion_RM$partes_por_billion ~ df_estacion_RM$PA_mmHg +
  df_estacion_RM$TMP_Celsius + df_estacion_RM$RH_porcentaje +
  df_estacion_RM$Direccion_viento + df_estacion_RM$Velocidad_viento,
  data = df_estacion_RM)

Residuals:
    1      2      3      4      5      6      7
-0.128586 -0.225894  0.481909 -0.261520 -0.011892  0.008344  0.137640

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    208.576774  377.642340   0.552   0.679
df_estacion_RM$PA_mmHg
-0.333629    0.643021  -0.519   0.695
df_estacion_RM$TMP_Celsius
-0.345155    0.319124  -1.082   0.475
df_estacion_RM$RH_porcentaje
-0.061127    0.034698  -1.762   0.329
df_estacion_RM$Direccion_viento
-0.006329    0.039450  -0.160   0.899
df_estacion_RM$Velocidad_viento
-0.174682    1.505583  -0.116   0.926

Residual standard error: 0.6224 on 1 degrees of freedom
Multiple R-squared:  0.9141,    Adjusted R-squared:  0.4849
F-statistic: 2.129 on 5 and 1 DF,  p-value: 0.4764

```

Figura 9. Resultados Regresión Múltiple estación Montecillo, Texcoco.

Gráfico de Regresión Múltiple:

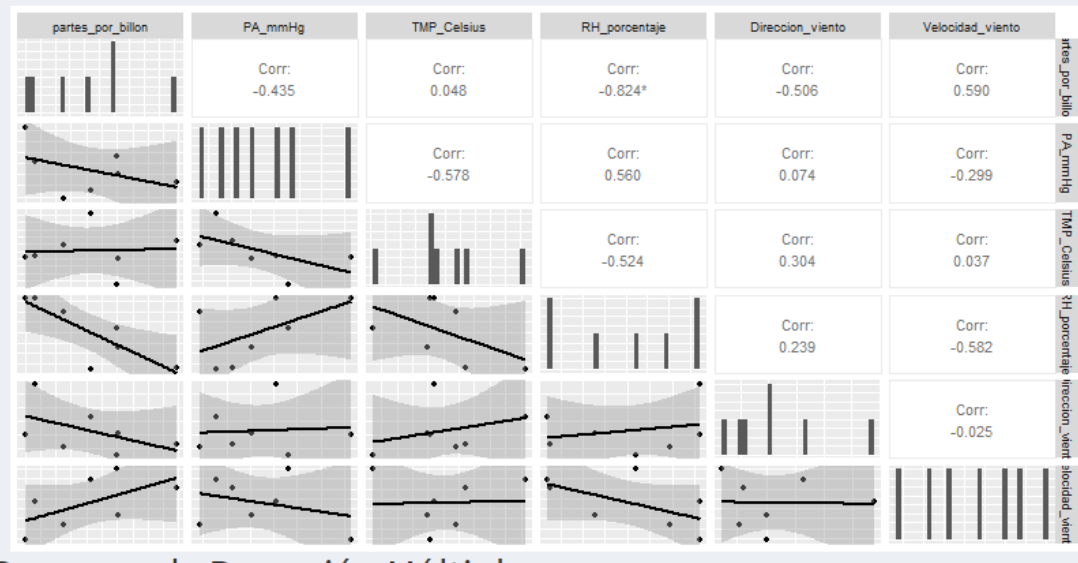


Figura 10. Gráfica regresión múltiple estación Montecillo, Texcoco.

Gráfico de Correlación:

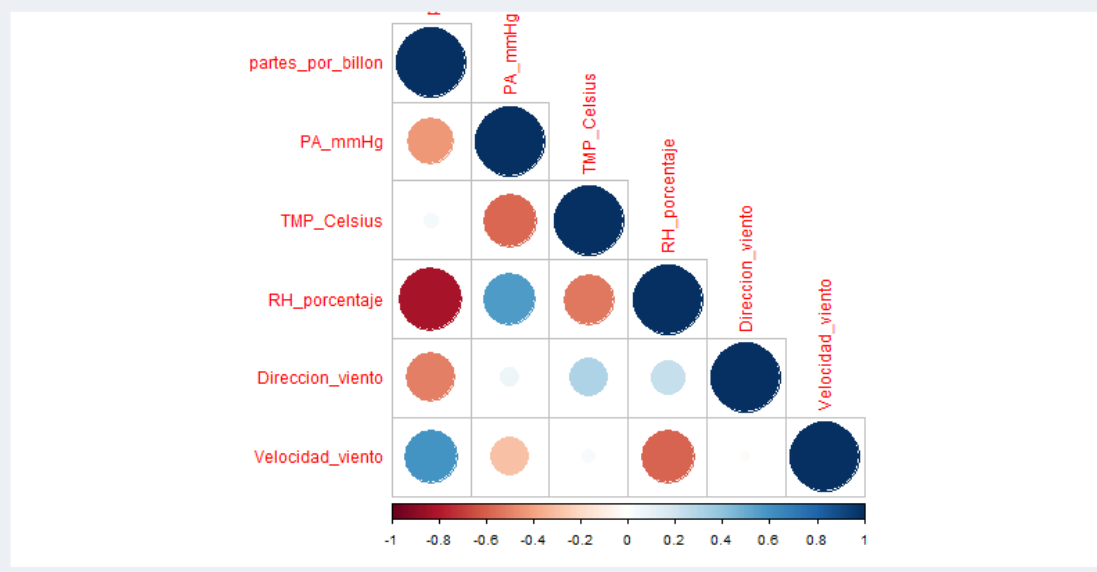


Figura 11. Gráfica correlación de la estación Montecillo, Texcoco.

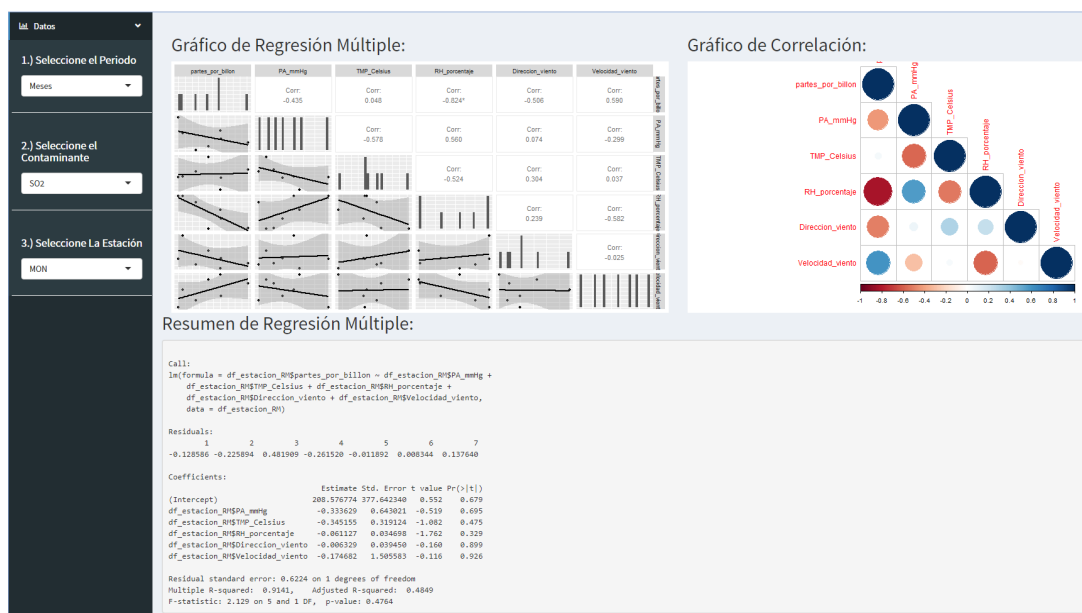


Figura 12. Ejemplo de la aplicación web local para la regresión lineal

c.- Resultados obtenidos.

Responsable: Tania Camila Castillo Robles

Gracias a las gráficas generadas en R y resultados obtenidos con nuestras regresiones tanto lineal como múltiple, pudimos observar que sí hay cierta correlación entre las variables meteorológicas y la concentración de los contaminantes en la atmósfera, aunque no sea tan fuerte. Además, también pudimos observar que se presentó una correlación más alta entre los RH porcentaje, la Dirección de viento y la Velocidad de viento con las partes por billón. También, gracias a la regresión lineal entre el dióxido de azufre y el tiempo realizado a partir de los últimos 10 años, pudimos observar por medio del coeficiente de determinación obtenido que se explica en Montecillo, Texcoco, Estado de México un 61% la relación que hay entre los datos examinados.

Estas variables afectan de manera no tan directa a la concentración de los contaminantes, sin embargo son importantemente visibles para el crecimiento de la concentración de contaminantes en la atmósfera, por lo que consideramos importante que futuros estudios tomen en cuenta este tipo de resultados para poder indagar y obtener más información al respecto de la correlación existente entre estas.

De manera más detallada, a partir de la regresión lineal se obtuvo que las concentraciones de dióxido de azufre tienen una correlación negativa respecto al tiempo, es decir, conforme el tiempo va avanzando, la cantidad de dióxido de azufre va disminuyendo. Este comportamiento para la estación seleccionada (Montecillo, Texcoco), se puede modelar mediante la siguiente ecuación:

$$y = - 0.0004053x + 8.662$$

Donde y es la concentración del dióxido de azufre en partículas por billón. y x es la fecha en formato numérico. Esta regresión explica de manera general el comportamiento de los datos, pues como se mencionó anteriormente, se tiene un valor de r^2 de 0.61, por lo que explica a grandes rasgos su

comportamiento pero no de manera exacta. Y aunque se podría usar un sistema no lineal, este podría ocasionar un sobreajuste, por lo que explicaría de manera exacta los datos, pero fallaría mucho para explicar los datos posteriores

En cuanto a la regresión múltiple, se obtuvieron diferentes valores de correlación entre las diferentes variables meteorológicas y la concentración de las partículas de dióxido de azufre. Entre estas variables meteorológicas se encontró que las que más se correlacionaron con este contaminante fueron el porcentaje de humedad, la dirección y velocidad del viento. Al final, se obtuvo un modelo matemático que estima la concentración de partículas de dióxido de azufre con base en la presión atmosférica (mmHG, x_1), la temperatura (grados Celsius, x_2), humedad relativa (porcentaje, x_3), dirección del viento (grados, x_4), velocidad del viento (metros por segundo, x_5):

$$y = -0.3336x_1 - 0.3452x_2 - 0.0611x_3 - 0.0063x_4 - 0.1746x_5 + 208.5767$$

Con base en esto, se obtuvo un modelo muy bueno, ya que el valor de r^2 fue de 0.9141, por lo que el modelo satisface muy bien los datos proporcionados.

4.- CIERRE:

a.- Síntesis. Presentar de manera creativa un resumen de lo hecho (aprendizajes adquiridos).

Responsable. Elisa Sánchez Bahnsen

A manera de resumen, para la solución de la situación problema, primero se encontraron las bases de datos a las cuales les podríamos aplicar los métodos estadísticos y así hallar una posible relación entre los factores meteorológicos y la contaminación. Los datos utilizados fueron proporcionados por el Sistema de Monitoreo Atmosférico de la Zona Metropolitana de la Ciudad de México. A partir de ellos, se eligió el software R para realizar una limpieza y ordenamiento de los mismos y posteriormente generar una regresión lineal y múltiple. Al generar estas dos regresiones pudimos descubrir ciertos patrones y tendencias que nos proporcionaron información acerca de aquellos factores que empeoran las condiciones ambientales relacionadas a la contaminación. No únicamente obtuvimos información sobre estas cuestiones ambientales, sino que también pudimos poner en práctica nuestra capacidad de análisis estadístico a través de todos los métodos aprendidos durante el curso. Se supieron calcular y graficar elementos de la estadística como la ecuación de la regresión, los coeficientes de determinación, correlación y los p-value. Además, los resultados no fueron únicamente obtenidos y presentados, sino que tuvimos la capacidad de interpretarlos correctamente para darles un significado estadístico que estuviera relacionado con la situación problema para después generar las conclusiones y observaciones adecuadas. Con esto último esperamos que lo encontrado pueda convertirse en información de valor que sirva posteriormente para proponer soluciones que combatan el problema de la contaminación.

Referencias:

- Hayas A. (1991) Meteorología y contaminación atmosférica. Peculiaridades de la Zona urbana de Jaén. Facultad de Ciencias Experimentales . Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=1202715>

- Coppel,E .(2017) Qué es la inversión térmica y cómo afecta a la contaminación de la Ciudad de México. El País. Recuperado de https://verne.elpais.com/verne/2017/12/06/mexico/1512516220_282144.html
- Dirección de Monitoreo Atmosférico. (s. f.). Gobierno de la ciudad de Mexico. Recuperado 9 de junio de 2021, de <http://www.aire.cdmx.gob.mx/default.php?opc=%27aKBhnmI=%27&opcion=bg==>
- Santos, D. (2017). Altas presiones que contaminan. Meteored. Recuperado de: <https://www.tiempo.com/noticias/ciencia/altas-presiones-que-contaminan.html>
- Dirección de Monitoreo Atmosférico. (2021). Gobierno de la Ciudad de México. Recuperado de <http://www.aire.cdmx.gob.mx/default.php?opc=%27ZaBhnmI=&dc=%27ZA==>