PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

Facultad de Matemáticas, Departamento de Estadística

**DIPLOMADO EN ESTADISTICA**

**Evaluación - Regresión Lineal**

29 de julio de 2024

El objetivo es entender y explicar el comportamiento de los niveles de contaminación del aire en la Región Metropolitana. Para ello, desde el Sistema de Información Nacional de Calidad del Aire (sinca.mma.gob.cl), se seleccionó una muestra de la información histórica de la Estación Parque O’Higgins en Santiago, la cual ha sido almacenada en la base ***Contam.xlsx.*** Se dispone de las siguientes variables:

* **PM2.5** – Materia particulada de 2.5 mg/m3. El PM2.5 son partículas muy pequeñas suspendidas en el aire que tienen un diámetro de menos de 2.5 micras. La materia particulada incluye sustancias químicas orgánicas, polvo, hollín y metales. Es nuestra variable respuesta.
* Potenciales variables explicativas:

|  |  |
| --- | --- |
| **Variables meteorológicas** | **Contaminantes atmosféricos** |
| Viento – Velocidad del viento (m/s) | NO – Monóxido de nitrógeno (ppb) |
| TProm – Temperatura promedio (° Celsius) | NO2 – Dióxido de nitrógeno (ppb) |
| TMin – Temperatura mínima (° Celsius) | CO – Monóxido de carbono (ppm) |
| TMax – Temperatura máxima (° Celsius) | O3 – Ozono (ppb) |
| Humed – Humedad relativa del aire (%) |  |

Usted, junto a su grupo, debe responder cada una de las siguientes interrogantes:

1. Obtenga el mejor modelo de regresión lineal simple basado en las variables meteorológicas.
2. Obtenga el mejor modelo de regresión lineal simple basado en los contaminantes atmosféricos.

Para los modelos simples:

* Puede recurrir a transformar las variables de ser necesario.
* Revise los supuestos y explique el modelo escogido en cuanto a sus resultados e interpretación de beta1).

1. Con base a todas las variables (meteorológicas y contaminantes), mediante una técnica iterativa (forward o backward) seleccione el mejor modelo predictivo. Indique para cada paso qué variable entra/sale del modelo, indicando el aumento/disminución del R2-ajustado.
2. Basado en los resultados previos, proponga un modelo con tres predictores (debe incluir una variable meteorológica y dos contaminantes), revise supuestos y evalúe con especial énfasis el problema de multicolinealidad. Apóyese de tablas de correlación, gráficos y métricas respectivas.

**Indicaciones**:

* Redacte un informe (pregunta/respuesta) que permita responder cada punto indicado.
* NO coloque el script en el informe (solo tablas, gráficos, métricas de interés, valores resultantes que ayuden a respaldar sus respuestas, no muy relleno).
* El trabajo puede ser grupal (hasta 2 personas, sin más excepciones).
* El informe se entrega por buzón (disponible en la plataforma del curso).
* **Fecha de entrega: 31 de julio hasta las 18:00 hrs.**

Referencia: <https://sinca.mma.gob.cl/index.php/estacion/index/id/273>