

Programación Actuarial III

Caso: Contaminación Ambiental

Introducción

Para esta actividad tendrás que escribir 3 funciones que interactúen con la base de datos que se anexa `specdata.zip`

Datos

El archivo se puede descargar de:

<https://www.dropbox.com/s/2ds1qwsv4uf9w5r/specdata.zip?dl=0>

El archivo contiene 332 archivos de valores separados por coma (CSV) que contienen los datos de un monitoreo para contaminantes del aire como partículas de materias finas (PM) en 322 ciudades. Cada registro tiene información de un monitor y el número de ID de cada monitor está registrado en el nombre del archivo. Por ejemplo, los datos del monitor 200 están contenidos en el archivo "200.csv". Cada archivo tiene 3 variables:

- Date: la fecha de la observación en formato AAAA-MM-DD
- Sulfate: el nivel de PMs de sulfato en esa fecha ($\mu\text{g}/\text{m}^3$)
- Nitrate: el nivel de PMs de nitrato en esa fecha ($\mu\text{g}/\text{m}^3$)

Para esta actividad deberás descargar la información en tu directorio de trabajo y crear un directorio llamado "specdata". No modifiques los archivos, notarás que hay algunos valores faltantes para diversas fechas de sulfatos o nitratos, esto es muy normal.

Parte 1

Escribe una función llamada "mediacontaminante" que calcule la media de algún contaminante (sulfato o nitrato) a lo largo de una lista de monitores. La función tomará tres argumentos "directorio", "contaminante", "id". Dado un vector de números de ID de monitores, "mediacontaminante" leerá los datos de contaminantes de ese monitor en el archivo que se encuentre contenido en el directorio especificado en el argumento "directorio" y devolverá como resultado la media del contaminante a lo largo de los monitores ignorando los valores faltantes. El esqueleto de la función debe quedar así:

```
mediacontaminante <- function(directorio, contaminante, id = 1:332) {  
  ## 'directorio' es un vector de caracteres de tamaño 1  
  ## con la dirección de la ubicación de las bases de datos  
  
  ## 'contaminante' es un vector de caracteres de tamaño 1 que indica  
  ## el nombre del contaminante que se determinará la media; ya sea  
  ## "sulfate" o "nitrate"
```

```
## 'id' es un vector de enteros que indica los ID's de los monitores  
## a usar.  
  
## Regresa la media del contaminante que se solicita a lo largo de  
## los monitores mencionados en el vector. (ignorando los NA's)  
}
```

Salva tu código en un script llamado “**mediacontaminante.R**”.

Parte 2

Escribir una función que lea el directorio lleno de archivos y reporte los números de registros completos en cada archivo. La función deberá regresar un data frame en el que la primera columna sea el ID del monitor y la segunda columna, el número de casos completos. El esqueleto de la función debe ser como sigue:

```
completos <- function(directorio, id = 1:332) {  
  ## 'directorio' es un vector de caracteres que indica la  
  ## ubicación de los archivos CSV.  
  
  ## 'id' es un vector de enteros indicando los monitores a  
  ## ser utilizados en el análisis.  
  
  ## El resultado es un data frame que se ve de esta manera:  
  ## id nobs  
  ## 1 117  
  ## 2 1041  
  ## ...  
  ## donde 'id' es el ID del monitor y 'nobs' es el número de  
  ## casos completos en el archivo de ese monitor.  
}
```

Salva tu código en un script llamado “**completos.R**”.

Parte 3

Escribir una función que tome los datos de un directorio de datos y un horizonte de casos completos y que con ellos calcule la correlación entre sulfato y nitrato para las localidades en donde el número de casos completos observados (en todas las variables) es mayor que el horizonte. La función debe devolver un vector de correlaciones para los monitores que cumplen con el horizonte requerido (cantidad mínima de casos completos). Si ningún monitor cumple el requerimiento de horizonte, la función debe devolver un vector numérico de tamaño 0. La función debería verse así:

```
corr <- function(directorio, horizonte = 0) {  
  ## 'directorio' es un vector de caracteres de tamaño 1 con la  
  ## dirección de ubicación de los archivos.  
}
```

```
## 'horizonte' es un vector numérico de tamaño 1 que indica el
## número de casos completos requeridos para calcular la
## correlación entre nitrato y sulfato en ese monitor.

## Regresa un vector con las correlaciones de los que cumplieron
}
```

Salva tu código en un script llamado “**corr.R**”.