

PAC2 Primavera 2023

UOC

Introducción

Las PECS se basarán en una base de datos obtenida a partir del repositorio de microdades del “Banco Mundial” a <https://microdata.worldbank.org/index.php/catalog/424/get-microdata>

Contiene indicaciones, entre otras de

1. *City* = Nombre de la ciudad
2. *Country* = País
3. *Population2000* = Población de la ciudad en 2000.
4. *PM10Concentration1999* = *PM10 concentrations (micro gramos por cubic meter) in residential areas of cities larger than 100,000*, en 1999
5. *Region* = Clasificación en región geográfica
6. *IncomeGroup* = Clasificación según nivel de ingresos del país.

Para importar los datos podemos usar la siguiente instrucción:

```
datosPM10<-read.table("../AirPollution2000WB_UOC2.csv", header=TRUE,
                        sep=";",na.strings="NA",
                        fileEncoding = "UTF-8", quote = "\"",
                        colClasses=c(rep("character",4),rep("numeric",2),
                                    rep("character",2)))
```

Hay que entregar la práctica en formato pdf (exportando el resultado final a pdf por ejemplo) en esta misma tarea Moodle; no hay que entregarla en el registro de EC.

Os puede ser útil la función `table` para tabular los datos.

Mirar las actividades resueltas de probabilidad del reto 2.

NOMBRE:

PAC2

Una vez importados los datos

Problema 1 (30 puntos)

- Escogemos una ciudad al azar. Calcular la probabilidad de que esta ciudad tenga una concentración de PM10 superior a la media. (15 puntos)
- Escogemos una ciudad de Norteamérica. Calcular la probabilidad de que esta ciudad tenga una concentración de PM10 superior a la media de las ciudades de América del Norte y que además tenga una población superior a los 50000 habitantes en el año 2000. (15 puntos)

Solución

- Para calcular la probabilidad que esta ciudad tenga una concentración de PM10 superior a la media, primero calculamos cuántas ciudades hay con una concentración de PM10 superior a la media:

```
(numero.ciudades = length(datosPM10$PM10Concentration1999  
                           [datosPM10$PM10Concentration1999 >  
                             mean(datosPM10$PM10Concentration1999)]))
```

```
## [1] 1157
```

A continuación, calculamos el número de ciudades totales:

```
(numero.ciudades.totales = dim(datosPM10)[1])
```

```
## [1] 3218
```

La probabilidad será, pues,

```
(p1=numero.ciudades/numero.ciudades.totales)
```

```
## [1] 0.3595401
```

- Creemos primero la subtabla de ciudades de América del Norte:

```
datosPM10.AmericaNorte = datosPM10[datosPM10$Region=="North America",]
```

Para calcular la probabilidad de que esta ciudad tenga una concentración de PM10 superior a la media y con una población superior a 50000 habitantes, primero calculamos cuántas ciudades hay con una concentración de PM10 superior a la media y con una población superior a 50000 habitantes:

```
(numero.ciudades.AmericaNorte =  
  length(datosPM10.AmericaNorte$PM10Concentration1999  
    [datosPM10.AmericaNorte$PM10Concentration1999 >  
      mean(datosPM10.AmericaNorte$PM10Concentration1999)  
      & datosPM10.AmericaNorte$Population2000 > 50000]))
```

```
## [1] 99
```

A continuación, calculamos el número de ciudades totales de América del Norte:

```
(numero.ciudades.totales.AmericaNorte = dim(datosPM10.AmericaNorte)[1])
```

```
## [1] 256
```

La probabilidad será, pues,

```
(p1=numero.ciudades.AmericaNorte/numero.ciudades.totales.AmericaNorte)
```

```
## [1] 0.3867188
```

Problema 2 (70 puntos)

- Crear la tabla que da el número de ciudades según región y nivel de ingresos. (15 puntos)
- Calcular la probabilidad de que una ciudad pertenezca a la región “Este de Asia y Pacífico” y tenga un nivel de ingresos “Lower middle income”. (10 puntos)
- Calcular la probabilidad de que una ciudad de la región “América Latina y Caribe” tenga un nivel de ingresos bajo. (15 puntos)
- Calcular la probabilidad de que una ciudad con un nivel de ingresos alto pertenezca a la región “Europa y Asia Central”. (15 puntos)
- Los sucesos “Pertener a la región Europa y Asia Central” y “Tener un nivel de ingresos alto”, son independientes? Razonar la respuesta. (15 puntos)

Solución

- La tabla es la siguiente:

```
(tabla=table(datosPM10$Region,datosPM10$IncomeGroup))
```

```
##
##               High income Low income Lower middle income
## East Asia & Pacific          284           7             143
## Europe & Central Asia        487           2              72
## Latin America & Caribbean     38           3              21
## Middle East & North Africa     30          12              46
## North America                256           0               0
## South Asia                    0           6            388
## Sub-Saharan Africa            0          69             86
##
##               Upper middle income
## East Asia & Pacific              405
## Europe & Central Asia            310
## Latin America & Caribbean        405
## Middle East & North Africa        103
## North America                    0
## South Asia                       8
## Sub-Saharan Africa               37
```

b) La probabilidad será:

$$pb = \frac{143}{3218} = 0.0444375.$$

c) La probabilidad será:

$$pc = \frac{3}{467} = 0.006424.$$

d) La probabilidad será:

$$pc = \frac{487}{1095} = 0.4447489.$$

e) La probabilidad del suceso “Ser de Europa y de Asia Central” vale:

$$p.Europa.Asia.Central = \frac{871}{3218} = 0.270665.$$

La probabilidad del suceso “Tener un nivel de ingresos alto” vale:

$$p.Ingresos.Alto = \frac{1095}{3218} = 0.3402735.$$

La probabilidad de la intersección de los dos acontecimientos anteriores vale:

$$p.inter = \frac{487}{3218} = 0.1513362.$$

Como la probabilidad de la intersección no es el producto de las probabilidades, los sucesos no son independientes:

$$\begin{aligned} p.inter &= 0.1513362 \neq p.Europa.Asia.Central \cdot p.Ingresos.Alto \\ &= 0.270665 \cdot 0.3402735 = 0.0921001. \end{aligned}$$