

## PEC1 primavera 2023

### UOC

Las PECs se basarán en una base de datos obtenida a partir del repositorio de microdatos del “Banco Mundial” en

<https://microdata.worldbank.org/index.php/catalog/424/get-microdata>

Contiene indicaciones, entre otros 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:

```
dadesPM10<-read.table("C:/Users/Paco/Desktop/documentos_paco/uoc/uoc_2021/AirPollution2000WB_UOC2.CSV", header=TRUE,
  sep=";", na.strings="NA",
  fileEncoding = "UTF-8", quote = "\"",
  colClasses=c(rep("character",4),rep("numeric",2),
    rep("character",2)))
```

Os puede ser útil consultar el siguiente material del reto 1:

- El entorno estadístico R. Estructura, lenguaje y sintaxis
- Análisis de datos y estadística descriptiva con R
- Actividades Resueltas del Reto 1 (Estadística Descriptiva)

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

**NOMBRE:**

### PEC1

Una vez importados los datos

### Pregunta 1. (25%) Sobre la variable `PM10Concentration1999` encontrad:

- a) Los resúmenes numéricos. (5%)
- b) La desviación estándar. (5%)
- c) El histograma y los resúmenes numéricos de la variable `PM10Concentration1999` en las ciudades de Algeria y Russian Federation. Comentad los resultados comparando ambos países y a la vez con los resultados globales del total de ciudades obtenidos en el apartado a). (15%)

Solución:

#a)

```
summary(dadesPM10$PM10Concentration1999)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      6.0    24.0    38.0    51.1    71.0   359.0
```

#b)

```
sd(dadesPM10$PM10Concentration1999)
```

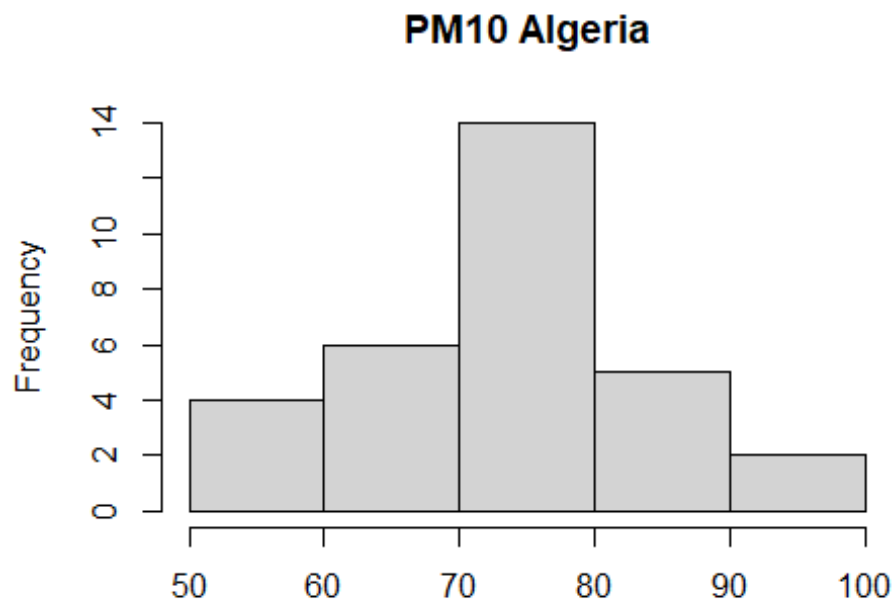
```
## [1] 38.14492
```

#c)

```
summary(dadesPM10$PM10Concentration1999[dadesPM10$Country=="Algeria"])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    50.00  68.50   73.00   73.52  79.50   98.00
```

```
hist(dadesPM10$PM10Concentration1999[dadesPM10$Country=="Algeria"],main="
PM10 Algeria")
```

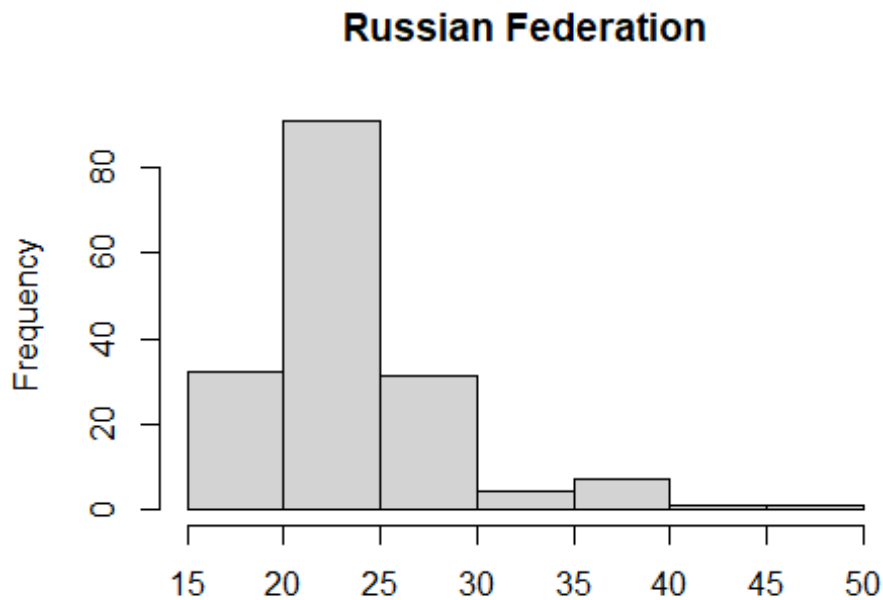


```
dadesPM10$PM10Concentration1999[dadesPM10$Country == "Alg
```

```
summary(dadesPM10$PM10Concentration1999[dadesPM10$Country=="Russian Feder  
ation"])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##   15.00   21.00   23.00   23.87   26.00   47.00
```

```
hist(dadesPM10$PM10Concentration1999[dadesPM10$Country=="Russian Federati  
on"],main="Russian Federation")
```



```
PM10$PM10Concentration1999[dadesPM10$Country == "Russian F
```

En las ciudades de Algeria tenemos bastante simetría y es unimodal. La concentración que tienen las ciudades está comprendida entre 50 y 100 micro gramos por metro cúbico. La media de las ciudades de Algeria (73.52) está por encima de la media global vista en el apartado a)(51.1).

En las de la Federación Rusa tenemos asimetría con cola hacia la derecha; la media está desplazada respecto a la mediana. La mayoría de las ciudades estudiadas tienen una concentración entre 15 y 30 micro gramos por metro cúbico. Hay muy pocas ciudades con una concentración mayor de 30. La media (23.87) está por debajo de la media global vista en el apartado a)(51.1)

**Pregunta 2. (20%) Encontrad el valor máximo de la variable PM10Concentration1999 con la ciudad y el correspondiente país donde se da este valor, así como el grupo de ingresos donde se da este valor. (20%)**

Solución:

```
dadesPM10[which.max(dadesPM10$PM10Concentration1999),
c("City", "Country", "PM10Concentration1999", "IncomeGroup")]

##      City Country PM10Concentration1999      IncomeGroup
## 2668 Nyala   Sudan                   359 Lower middle income
```

**Pregunta 3. (30%)** Encontrad una tabla de frecuencias absolutas de la variable `IncomeGroup`, y otra tabla donde se muestren los porcentajes de cada grupo de ingresos (podéis usar `prop.table`). Haced un gráfico adecuado (de la frecuencia o del porcentaje), con las etiquetas verticales y ampliando el área del gráfico para que quepan (podéis utilizar la instrucción `par(mar=c(12,5,1,1))` y la opción `las=2`). Comentad los resultados.

Solución:

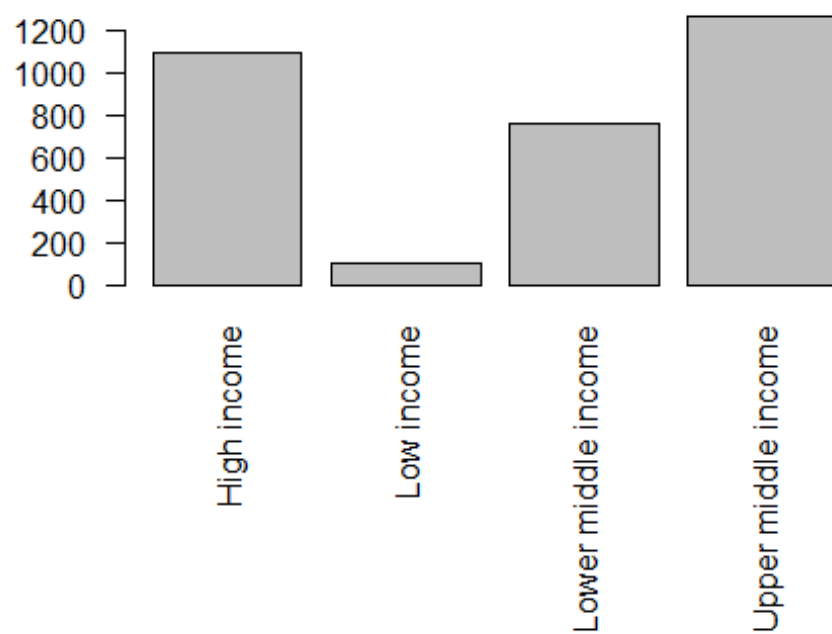
```
table(dadesPM10$IncomeGroup)

##
##           High income           Low income Lower middle income Upper midd
le income
##                1095                99                756
1268

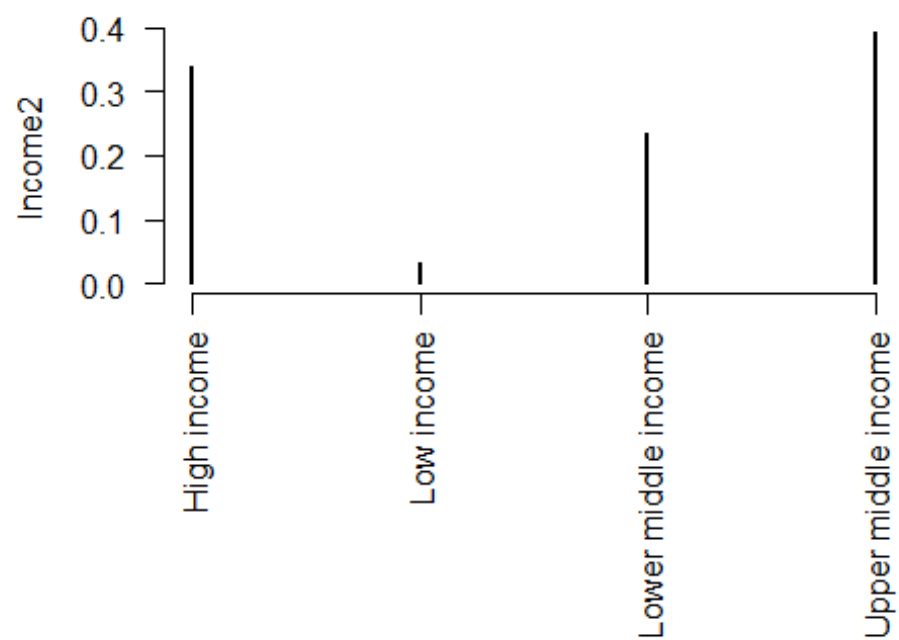
prop.table(table(dadesPM10$IncomeGroup))

##
##           High income           Low income Lower middle income Upper midd
le income
##           0.34027346           0.03076445           0.23492853           0
.39403356

Income<- table (dadesPM10$IncomeGroup)
Income2<- prop.table(table(dadesPM10$IncomeGroup))
par(mar=c(12,5,1,1))
barplot(Income,las=2)
```



```
plot(Income2, las=2)
```

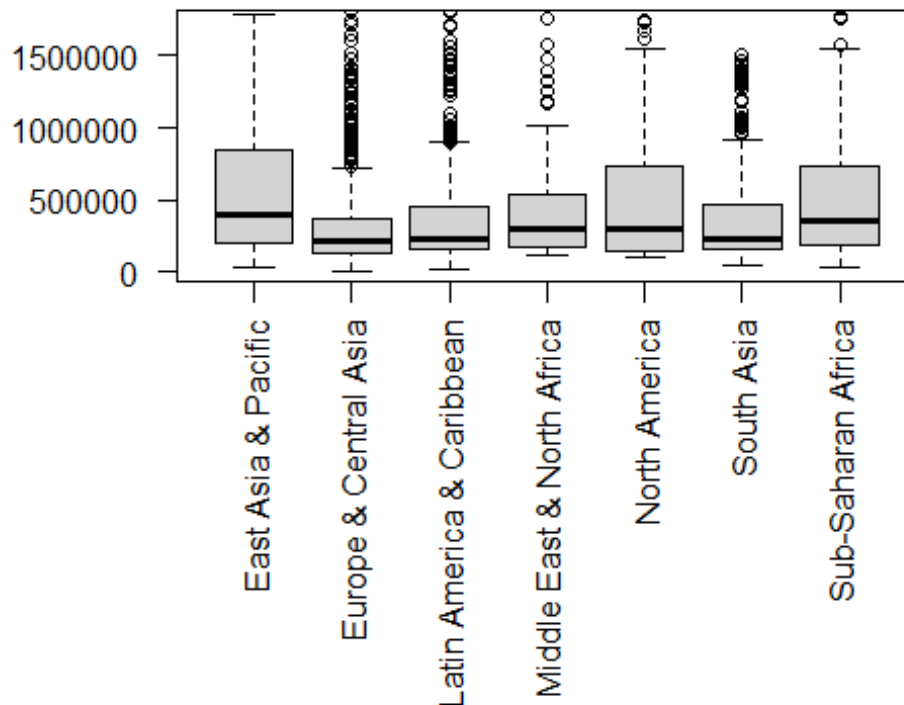


Hay más ciudades, en el estudio, con ingresos medio alto y altos. De las que menos hay es de ingresos bajos.

**Pregunta 4. (25%) Haced un Boxplot de Population2000 distinguiendo por regiones geográficas, con las etiquetas verticales. Los límites del eje vertical tienen que estar entre 0 y 1750000 y hay que ampliar el área del gráfico, para que se vean las cajas bien, con la instrucción `par(mar=c(12,5,1,1))`. Comentad los resultados.**

Solución:

```
par(mar=c(12,5,1,1))
boxplot(dadesPM10$Population2000~dadesPM10$Region,ylim=c(0.0e+00,1.75e+06),las=2,
        ylab="",xlab="")
```



Europa y Asia Central tienen valores más bajos y menos dispersos. En el otro extremo está el Pacífico y el Este de Asia con valores más altos y más dispersos.