

Tarea 1 - Análisis de Datos I

Enzo Loiza B.

9 de abril de 2024

Pregunta 1

Se realizó un análisis exploratorio a los siguientes datos de ciudades europeas y americanas:

Ciudad	Tamaño	Ingreso (USD)	Índice de Criminalidad	Calidad de Aire	Nivel de Desarrollo
París	Grande	60000	65	6.8	Alto
Zurich	Mediana	75000	30	8.0	Alto
Estambul	Grande	38000	58	6.7	Medio
Montevideo	Pequeña	40000	30	7.0	Bajo
Barcelona	Mediana	58000	35	7.5	Medio
Oslo	Mediana	72000	25	7.8	Alto
Londres	Grande	58000	30	7.0	Alto
Santiago	Grande	45000	38	6.3	Medio

a) En base a los datos, se crearon vectores para cada una de las columnas de manera manual asignando nombres correspondientes (consultar script). Luego, con estos datos, se creó una `data.frame` con estos vectores:

```
df <- data.frame(Ciudad = Ciudad,  
                  Tamaño = Tamano,  
                  "Ingreso USD" = Ingreso,  
                  "Índice Criminalidad" = Criminal_index,  
                  "Calidad Aire" = Air_quality,  
                  "Nivel Desarrollo" = Development_lvl)
```

Visualizamos el resultado mediante el comando `head()`, obteniendo el siguiente resultado:

	Ciudad	Tamaño	Ingreso.USD	Indice.Criminalidad	Calidad.Aire	Nivel.Desarrollo
1	Paris	Grande	60000	65	6.8	Alto
2	Zurich	Mediana	75000	30	8.0	Alto
3	Estambul	Grande	38000	58	6.7	Medio
4	Montevideo	Pequeña	40000	30	7.0	Bajo

b) Reportamos las medidas de tendencia central de las variables `Ingreso.USD`, `Indice.Criminalidad` y `Calidad.Aire`, obteniendo los siguientes valores de tendencia central:

Medida	Min.	1er Cuartil	Mediana	Moda	Media	3er Cuartil	Max.
Ingreso (USD)	38000	43750	38000	58000	55750	63000	75000
Índice de Criminalidad	25	30	32.50	30	38.88	43	65
Calidad de Aire	6.30	6.78	7.00	7.00	7.14	7.58	8.00

Y de dispersión:

Medida	Varianza	Desv. Estándar
Ingreso (USD)	191,642,857	13,843.51
Índice de Criminalidad	213.67	14.60
Calidad de Aire	0.34	0.58

c) Luego, generamos la tabla de frecuencias relativas para las variables **Tamaño**

Grande	Mediana	Pequeña
50.00	37.50	12.50

y **Nivel.Desarrollo**:

Alto	Medio	Bajo
50.00	37.50	12.50

Pregunta 2

a) Realizamos el gráfico de dispersión de ingreso y criminalidad. El código es similar en ambos casos, pues sólo cambiamos la fuente de unos ejes.

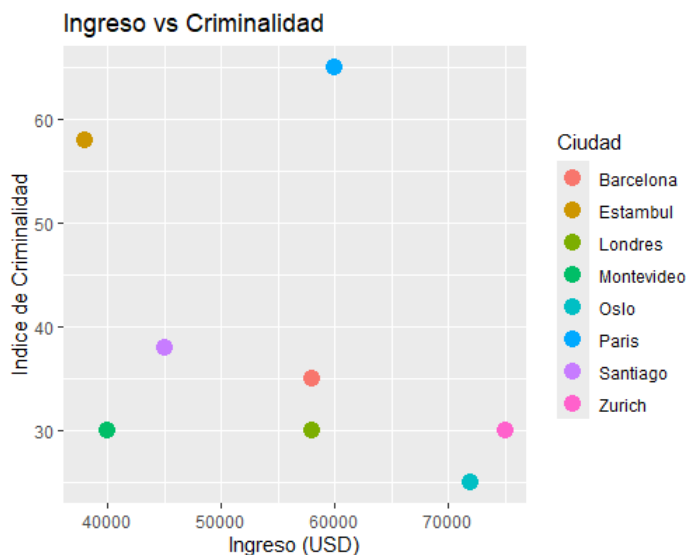


Figura 1: Ingreso (USD) versus Índice de Criminalidad

La figura 1, junto con un **coeficiente de correlación de -0.36**, indican que existe una correlación negativa moderada entre ambas variables. Esto es, a medida que aumenta el ingreso promedio en una región, tiende a haber una disminución en el índice de criminalidad, y viceversa. Es importante mencionar que esta correlación no indica causalidad, pues aunque puede que existe una influencia de uno sobre otro, puede haber otros factores que no se están considerando en el análisis.

b) Realizamos un análisis similar con las variables de Ingreso y Calidad de Aire.

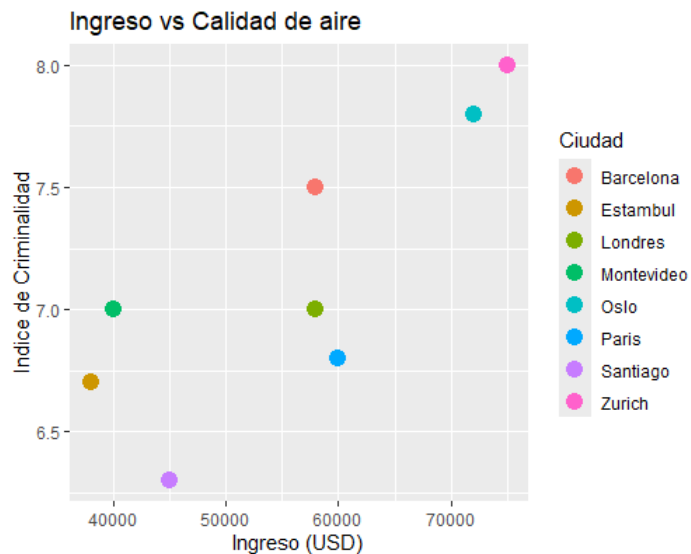


Figura 2: Ingreso (USD) versus Calidad de Aire

En este caso, en la Figura 2 se observa una relación fuerte entre el ingreso y la calidad de aire. Esto lo avala un **coeficiente de correlación de 0.81**, por lo que existe una correlación fuerte positiva entre ambas variables. A medida que aumenta el ingreso de una ciudad, tiende a haber una mejora en la calidad del aire. Sin embargo, al igual que con la correlación entre el ingreso y la criminalidad, es importante considerar varios factores adicionales para comprender completamente la relación^{1 2}.

¹Compilado en L^AT_EX

²Código disponible en GitHub