



Universidad Simón Bolívar  
Departamento de Cómputo Científico y Estadística

Estadística para ingenieros (CO3321)

## Laboratorio 1: Estadística descriptiva.

### 1. DATOS

Existe el interés en investigar la criminalidad en varios estados de los Estados Unidos. Para ello se dispone de datos de 47 estados relativos al año 1960, e incluye las siguientes variables.

- R Índice de criminalidad, número de delitos conocidos por la policía por cada millón de habitantes.
- Age Distribución de la edad, número de varones de edad 14-24 por cada mil de toda la población del estado.
- s Variable binaria que distingue entre estados del sur ( $S = 1$ ) del resto.
- Ed Nivel educativo, número medio de años de escolarización.
- $Ex_1$  Gasto per cápita en protección policial relativa a 1960.
- $Ex_2$  Gasto per cápita en protección policial relativa a 1959.
- LF Proporción en participación en trabajos de fuerza por cada mil hombres con edad 14-24.
- M Número de varones por mil mujeres.
- N Tamaño de la población del estado en cien miles.
- NW El número de personas de raza no blanca por mil habitantes.
- $U_1$  Razón de desempleo entre hombres de edad 14-24, por cada mil.
- $U_2$  Razón de desempleo entre hombres de edad 35-39, por cada mil.
- W Riqueza medida por el ingreso familiar.
- x Desigualdad en ingresos, el número de familias por mil que ganan por debajo de la mitad de la mediana de ingresos.

Estos se incluyen en el archivo `IndiceCrim.dat`.

### 2. EJERCICIOS

1. Realice un análisis descriptivo de las variables incluyendo: media muestral, mediana, cuartiles, máximo, mínimo, varianza, desviación estándar, rango y rango intercuartil. Interprete los resultados.
2. Realice un histograma y un diagrama de caja para cada variable. Interprete los gráficos.
3. Separe los datos dependiendo si los estados están en el sur o no. ¿Considera que hay diferencias por la ubicación geográfica del estado?

### 3. CONDICIONES DE ENTREGA

a Las variables con las que trabajarán serán:

1. Contreras, Di Battista, Mazutiel, Mendez, Münch, Prado y Sánchez. Variables R, Age, M.
2. Blanco, Macari, Medina, Pérez, Rodríguez, Salerno y Sucre. Variables R, Ed, N.
3. Adames, Becerra, Chaurio, Franco, Gelabert, Goudet, Lagrange. Variables R,  $Ex_1$ ,  $Ex_2$ .
4. Calderón, Cipagauta, De Ponte, Márquez, Palacios, Roa y Zeait. Variables R, LF, X.
5. Hernandez, Marin, Mejías, Quintana, Strubinger, Torres y Viera. Variables R, M, W.

6. Abascal, Aparicio, Garcia, Marrero, Peñaranda, Romero y Vásquez. Variables R, N, Age.
  7. Alexander, Altuve, Alzuro, Báez, Centeno, Espinoza y Nieves. Variables R, NW, Ed.
  8. Arteaga, Bandes, Borrell, Camacho, Guerra, Lovera y Palma. Variables R,  $U_1$ ,  $U_2$ .
  9. Cámara, Deftereos, Martinez, Navarro, Ortiz, Pizzo y Rivero. Variables R, W, NW.
  10. Amaya, Fernandez, Freites, Gonzalez, Hung, Rivas y Rodrigues. Variables R, X, LF.
- b Se debe realizar de forma individual o por parejas. En el caso de las parejas, deben escoger cual de los dos grupos de variables escoger (en el caso de que les toquen distintos grupos de variables).
  - c El documento a entregar debe incluir las salidas del paquete estadístico R necesarias para la interpretación. Este documento no debe exceder las 10 páginas y debe estar en formato .pdf.
  - d El documento debe estar debidamente identificado con el nombre de quien haya realizado el trabajo.
  - e La entrega se realizará en digital al correo electrónico povallesgarcia@usb.ve a más tardar 12 horas después del aviso de publicación. El asunto del correo DEBE ser: “*Laboratorio 1. CO3321*”.