



Universidad Simón Bolívar  
Departamento de Cómputo Científico y Estadística  
Estadística Para Ingenieros CO-3321  
Prof. Desiree Villalta

### Informe de Laboratorio: Laboratorio 1. Estadística descriptiva

Estudiantes:

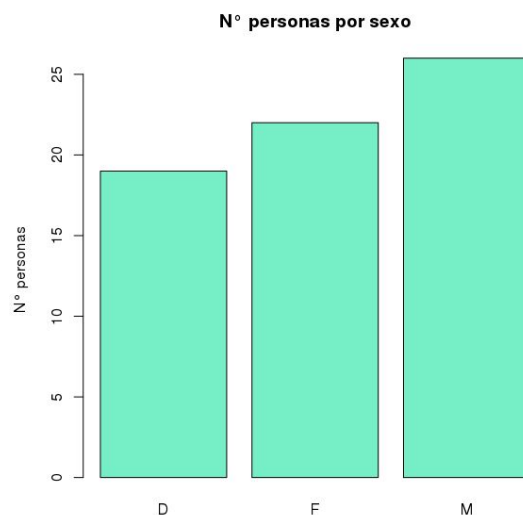
Carlos Sivira 15-11377

Jose Barrera 15-10123

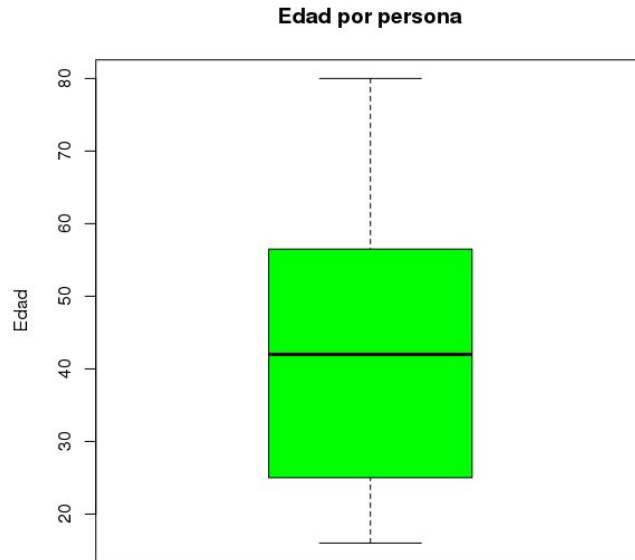
1. Realice un análisis descriptivo de las variables, apoyándose en los gráficos y el resumen de los datos.

La tabla suministrada presenta las siguientes variables con las siguientes características:

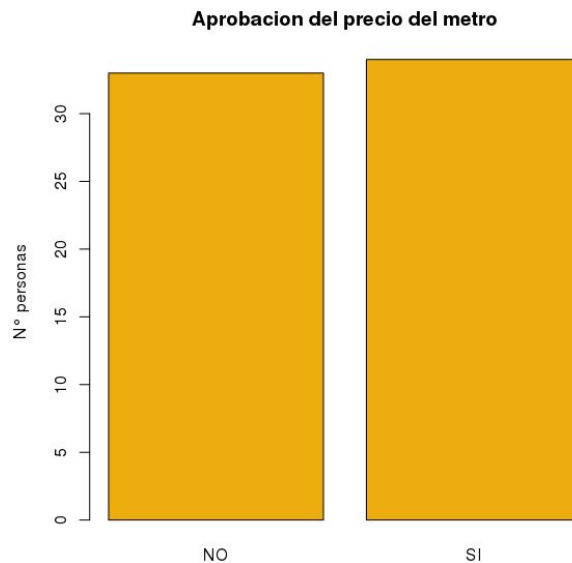
- Sexo: Variable cualitativa que muestra el sexo de una persona. Sus valores se distribuyen en: sexo diverso (D), femenino (F) y masculino (M). Existen 19 personas en D, 22 personas en F y 26 personas en M.



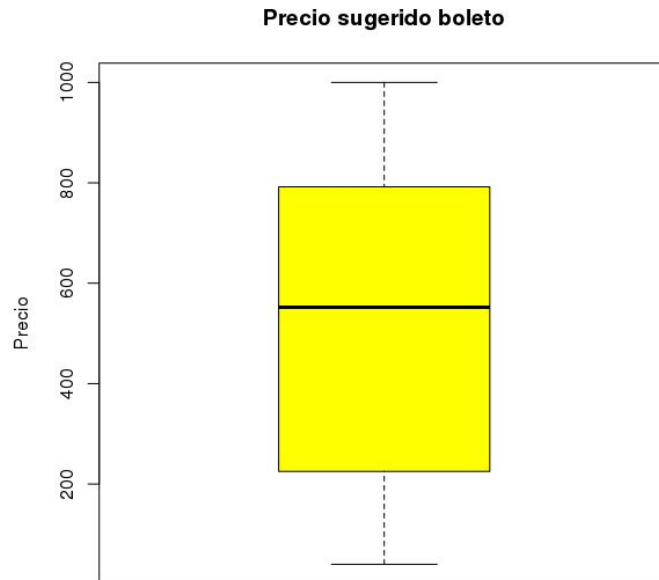
- Edad: Variable cualitativa que representa la edad de una persona. Sus valores varían entre 16 y 80, siendo su media 41.91 y su desviación estándar 18.31933. Los datos se encuentran menos dispersos, ya que el coeficiente de variación es un número aproximado a 0.4, lo que indica una distribución uniforme de los valores.



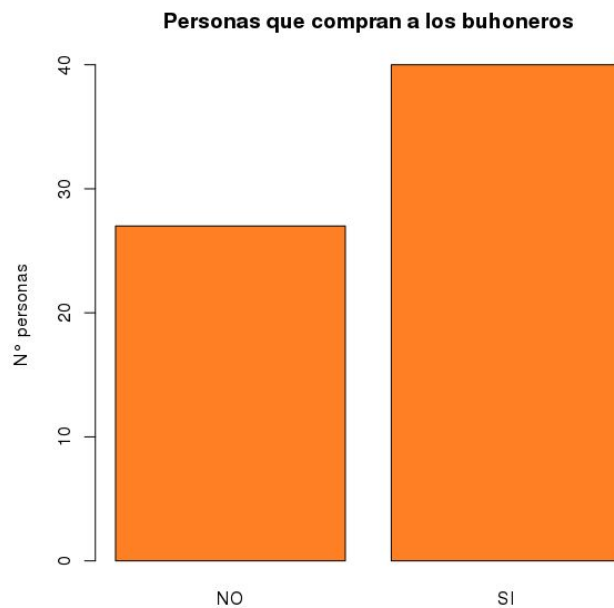
- Preg1: Variable cualitativa que muestra si una persona aprueba o no el precio del metro. La aprobación de la población es pareja, siendo 33 No y 34 Si.



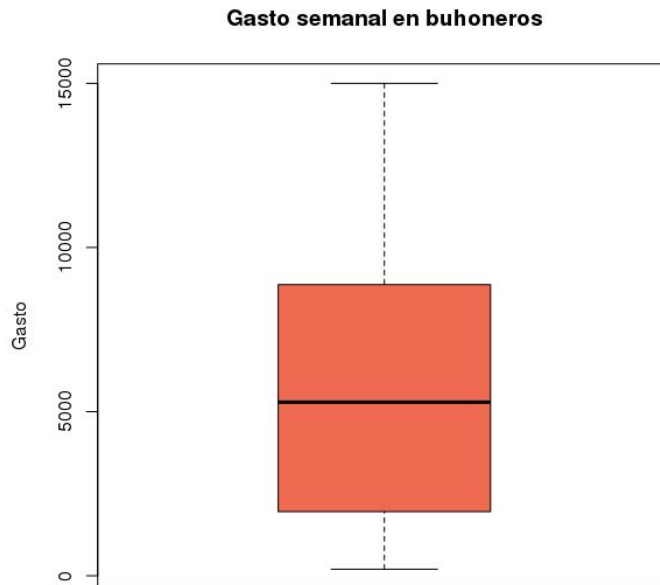
- Preg2: Variable cuantitativa que representa el costo que cada usuario está dispuesto a cancelar. Sus valores varían entre 400 y 1000, siendo su media 519.1 y su desviación estándar 322.4191. Al ver estos valores y compararlos con su gráfica, se evidencia que los datos se encuentran menos dispersos, ya que el coeficiente de variación es muy pequeño.



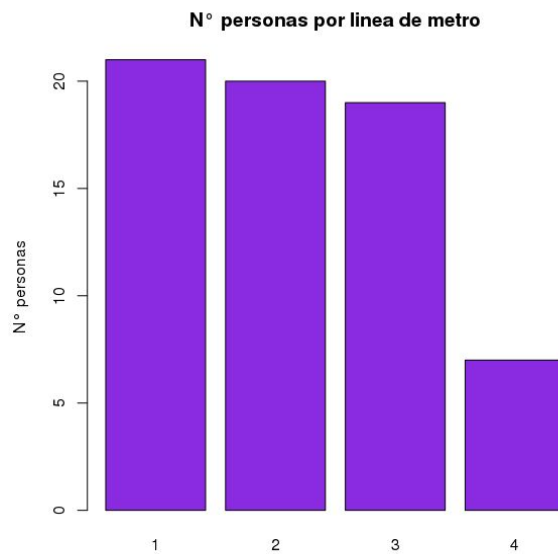
- **Preg3:** Variable cualitativa que muestra si un usuario compra boletos a vendedores ambulantes. 40 de los usuarios optan por comprar a buhoneros, mientras que 27 usuarios prefieren no hacerlo.



- **Preg4:** Variable cuantitativa que representa el gasto asociado a la compra de boletos a vendedores ambulantes. Sus valores varían entre 0 y 15000, siendo su media 3256 y su desviación estándar 3900.113. Estos valores evidencian una dispersión en los datos mucho mayor, donde la mayoría de los valores se encuentran entre 0 y 6417.

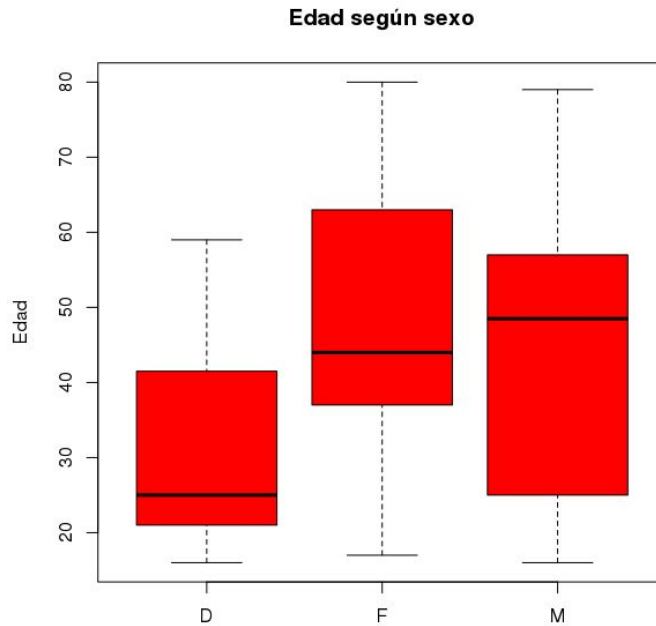


- Preg5: Variable cuantitativa que muestra qué línea de metro usa un usuario en particular. Los valores asociados varían entre la línea 1 a la 4 de forma discreta. La línea más usada es la 1, le siguen la 2 y 3 con valores muy similares. La línea 4 presenta un menor número de usuarios.



2. Para la muestra de personas encuestadas, ¿cómo se relaciona la edad según el sexo? En su análisis presente un diagrama de caja.

A continuación se muestra la distribución de edad, en el diagrama de caja de se visualiza la distribución por sexo.



Al aplicar la función summary sobre el conjunto de edades se obtienen los siguientes resultados:

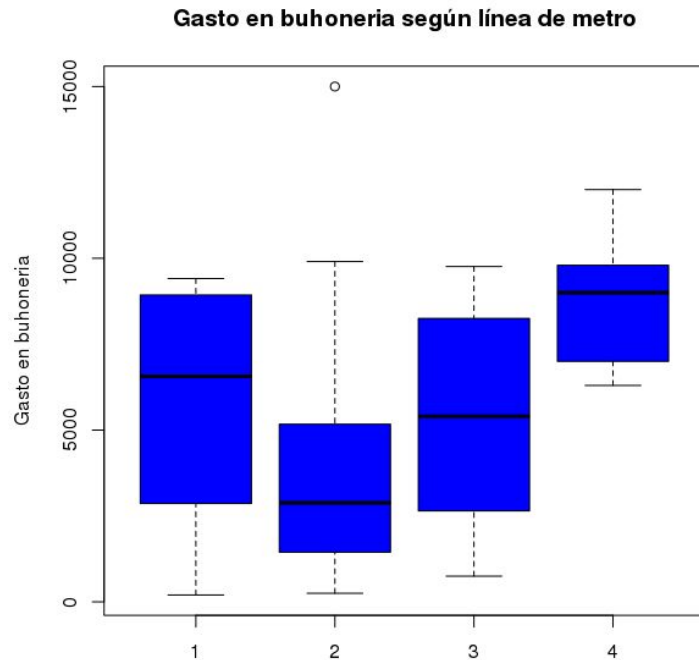
<b>Min.</b>	<b>1st Qu.</b>	<b>Median</b>	<b>Mean</b>	<b>3rd Qu.</b>	<b>Max.</b>	<b>Sd.</b>
16.00	25.00	42.00	41.91	56.50	80.00	18.31933

El resultado obtenido se ve reflejado en la gráfica anterior. En general, la población se encuentra muy bien distribuida en función de las edades (Gráfica: *Edad por persona*). Esto se sustenta en la separación existente entre la desviación estándar y la media de este conjunto de valores.

Sin embargo, al estudiar cada sexo, se evidencia que la distribución de valores es dispareja a lo largo de las edades. Caso claro es que el rango de Q1 a Q2 de sexo diverso (D) presenta una agrupación de valores en un rango muy pequeño, mientras que, el sexo masculino (M) presenta una mejor distribución a lo largo de todo su rango intercuartil. Así mismo, es de notar la inexistencia de valores atípicos en ambas gráficas, lo cual demuestra uniformidad en la distribución de las edades, sobre todo entre los sexos femeninos y masculinos.

3. De las personas que le compran a los vendedores ambulantes, ¿cuánto dinero gasta según la línea del metro en la cual desembarcan?. Apoye su análisis con un diagrama de caja.

A continuación se muestra la distribución de gasto en buhonería, en el diagrama de caja se visualiza la distribución por línea de metro usada.



Al aplicar la función summary sobre el conjunto de personas que realizan gasto en buhonería, se obtienen los siguientes resultados:

**Min. 1st Qu. Median Mean 3rd Qu. Max. Sd.**

0 0 1445 3256 6417 15000 3900.113

Nuevamente, existe una correspondencia entre los valores obtenidos y la gráfica mostrada. El rango de valores entre las líneas 1 y 3 es bastante similar, mientras que en la línea 4 se evidencia un rango distinto, donde un menor número de usuarios realiza un mayor gasto por uso. Así mismo, se denota la existencia de un dato atípico en la relación de gastos en la línea 2 que no se muestra en la gráfica de gastos de la población que compra a buhoneros (Gráfica: *Gasto semanal en buhonería*). Las líneas 1 y 3 presenta una distribución de gastos mucho más uniforme que en la líneas 2 y 4 donde los valores en el rango intercuartil son muy cercanos a la media calculada. Esto se visualiza en el coeficiente de variación, donde al dividir la desviación estándar entre la media se obtiene un valor aproximado a 1. Lo cual indica poca variación respecto a la media.