# **UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN**

**ESCUELA DE ESTADÍSTICA**

**PRIMER TRABAJO Análisis descriptivo de datos (10%) SEM. 02 – 2021**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # cédula usado |  |  | Grupo: |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre Completo** |  | **D. I.** |  |
| **Nombre Completo** |  | **D. I.** |  |

La base de datos con la que le corresponde trabajar se obtiene como una muestra aleatoria de una gran base de datos. La base original corresponde a la información de 465 estudiantes de la Universidad Nacional Sede Medellín. Dicha base contiene las variables: **EDAD** (en días), **GENERO** (HOMBRE o MUJER), **LONG**\_**ABDO** (longitud abdominal, en cm), **TIPO**\_**COLE** (PUBLICO, PRIVADO, OTROS), **LONG**\_**PIE** (Longitud promedio de los pies, en cm) y **LONG**\_**MAN** (Longitud promedio de las manos, en cm).

Cada pareja debe generar una base de datos, que corresponderá a una muestra de tamaño 100 de la base original. Los datos originales están en el archivo: “**base.txt**”, el cual está disponible en Moodle. Para hacerlo debe seguir los siguientes pasos:

# Primero leer el archivo de datos.

**uno <- read.table(file.choose(), header=T)**

# Copiar el siguiente código en R sin modificar nada

**genera <- function(cedula){**

**set.seed(cedula)**

**data <- uno[sample(1:465,100),]**

**data**

**}**

# Para crear la base de datos con la cual trabajara, debe ejecutar la siguiente línea:

**datos <- genera(cedula)**

“**cedula**” representa el número de cedula de uno de los dos integrantes. En el trabajo debe indicar cuál número de cédula usó. Este debe escribirse en la parte superior izquierda. En la parte superior derecha, el grupo al cual pertenecen los estudiantes. Se aclara que ambos estudiantes deben pertenecer al mismo grupo en el cual están matriculados.

Usando la base de datos generada, responda a las siguientes preguntas:

1. Para las variables **LONG**\_**ABDO** y **LONG**\_**MAN**, calcule: Media, Mediana y desviación estándar e interprete sus resultados en el contexto de la información. Para la variable **TIPO**\_**COLE**, elabore una tabla de porcentajes y el respectivo diagrama de barras. Comente los resultados obtenidos en el gráfico.
2. Si se considera la variable **EDAD**, ¿Esta se comporta igual en los hombres y mujeres? Justifique su respuesta. Elabore los gráficos que considere pertinentes. Estos deben ser interpretados, en el contexto de la información analizada.
3. ¿Se puede afirmar que los Hombres en promedio tienen los pies más largos que las Mujeres? Justifique su respuesta. Puede anexar los resúmenes y/o gráficos que considere pertinentes.
4. ¿Se puede afirmar que el porcentaje de Hombres que provienen de colegios públicos es mayor al porcentaje de Mujeres que vienen de colegios privados? Justifique su respuesta. Puede anexar los resúmenes y/o gráficos que considere pertinentes.

El trabajo debe cargarse en la carpeta definida en el curso en Moodle, en la pestaña ***Trabajos – Trabajo I – Soporte trabajo I***, antes de las **4:00 pm del sábado 23 de octubre de 2021**.