# **UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN**

**ESCUELA DE ESTADÍSTICA**

**PRIMER TRABAJO Análisis descriptivo de datos (10%) SEM. 01 – 2022**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # cédula usado |  |  | Profesor: |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **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 2100 personas colombianas. Dicha base contiene las variables: **SEXO** (Hom, Muj), **MASA** (en kg), **LONG\_ABDO** (Perímetro abdominal, en cm), **LONG**\_**PALMA** (Longitud promedio de las palmas de ambas manos, en cm) y **CAT\_IMC** (Categoría el IMC: Bajo, Medio, Alto).

Cada pareja debe generar una base de datos, que corresponderá a una muestra de tamaño **150** 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(1:2100,150),]**

**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. (**25 pts**.) Para las variables **MASA** y **LONG**\_**ABDO** calcule: Media, Mediana y desviación estándar e interprete sus resultados en el contexto de la información. Para la variable **CAT\_IMC**, elabore una tabla de porcentajes y el respectivo diagrama de barras. Comente los resultados obtenidos en el gráfico.
2. (**25 pts**.) Si se considera la variable **LONG**\_**ABDO**, ¿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. (**25 pts**.) ¿Se puede afirmar que los Hombres en promedio tienen las palmas de sus manos más largas que la de las Mujeres? Justifique su respuesta. Puede anexar los resúmenes y/o gráficos que considere pertinentes.
4. (**25 pts**.) ¿Se puede afirmar que el porcentaje de Hombres con un IMC alto es superior al porcentaje de Mujeres con un IMC alto? Justifique su respuesta. Puede anexar los resúmenes y/o gráficos que considere pertinentes.

El trabajo debe hacerse usando R o R-Commander.

Debe entregar un documento con la solución a las preguntas, **SOLO** en formato **pdf**. Este documento debe cargarse en el curso en Moodle, en la pestaña ***Trabajos – Trabajo I – Soporte Trabajo 1***, en la carpeta que le corresponde según su profesor y horario, antes de las **4:00 pm del sábado 26 de marzo de 2022**.