|  |  |
| --- | --- |
|  | **Universidad de Antioquia** **Facultad de Ingeniería**  **Departamento Ingeniería Industrial**  **Probabilidad e Inferencia Estadística**  **Semestre 2019 - II**  **Docente: Jorge Iván Pérez García** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre** |  | **Cédula** |  |
| **Nombre** |  | **Cédula** |  |
| **Semilla (Sumar de las dos Cédulas)** | | |  |

La base de datos contiene la información de personas que contrajeron el coronavirus y algunas características relacionadas con éstos. La base de datos contiene las siguientes variables.

* **País**: Hace referencia al país de origen de la persona que posee la enfermedad.
* **Sexo:** Hace referencia al sexo de la persona encuestada.
* **Peso:** Hace referencia al peso de la persona que posee coronavirus.
* **Altura:** Hace referencia a la altura de la persona que posee la enfermedad.
* **Edad:** Hace referencia a la edad de la persona que posee coronavirus.
* **Enf\_Resp:** Hace referencia a si la persona infectada con coronavirus, tiene además alguna enfermedad respiratoria.
* **Est\_Salud:** Hace referencia al estado de salud de la persona, tal que
  + **E:** Significa que la persona está actualmente enferma.
  + **R:** Significa que la persona se recuperó del coronavirus.
  + **M:** Significa que la persona falleció a causa de la enfermedad.

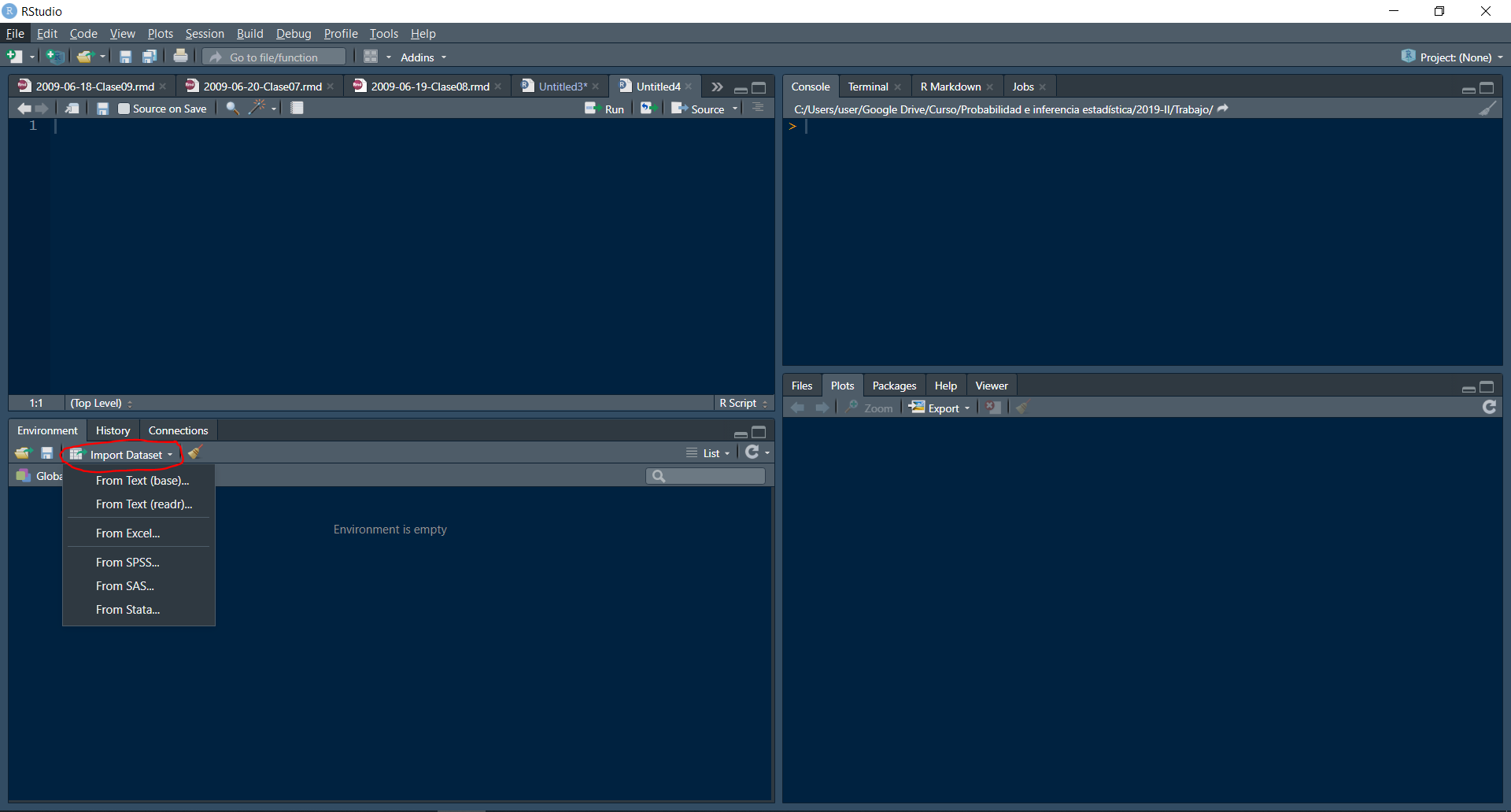
**LEA ATENTAMENTE:**

* Usted debe presentar un informe escrito de **MÁXIMO 5 PAGINAS**.
* Diligencie el nombre, la cédula y la semilla cuidadosamente.
* En caso de ser solo una persona, la semilla será igual a la cédula de la persona.
* La solución de las preguntas debe tener los resultados obtenidos en **R** y su debida interpretación junto a los gráficos que considere convenientes.
* Para facilitar la presentación de gráficos y salidas de **R**, emplee la **Herramienta de Recortes** de Windows**.**
* Una vez termine el trabajo deberá enviar el trabajo (documento de Word, y el código de R) al correo electrónico **jivan.perez@udea.edu.co**, con el nombre de los dos participantes en el asunto.

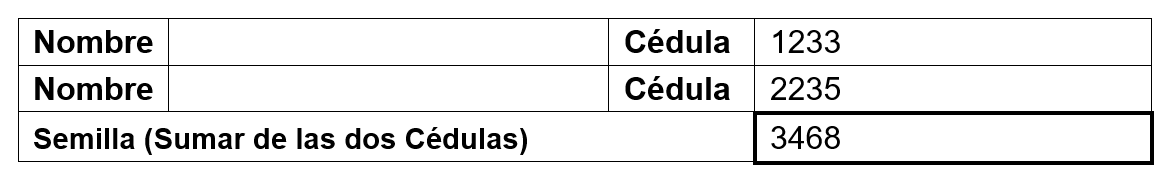
**ANTES DE INICIAR EL TRABAJO:**

Antes de iniciar con el desarrollo del trabajo siga los 3 siguientes pasos:

* Cargue la base de datos desde el formato .xlsx (From Excel…) o desde el formato .txt (From Text (base)…) desde la pestaña Import Dataset.



* Establezca una semilla en **R** igual al valor calculado al inicio de este documento



Mediante la siguiente línea de código de ejemplo (cambiar el número)

set.seed(**3468**)

* Haga la selección aleatoria de 250 datos, mediante el siguiente código (BD hace referencia al nombre con el cual cargaron la base de datos, cambie BD por nombre con el que haya cargado la base de datos)

datos <- **BD**[sample(1:nrow(**BD**), size = 250), ]

* Inicie con el desarrollo del trabajo. Si tienen dudas sobre cualquier código, función, librería o el trabajo como tal pueden escribirme al correo **jivan.perez@udea.edu.co**

**Preguntas**

Usando la información de su base de datos responda a las siguientes preguntas. **Justifique sus respuestas gráficamente y numéricamente, junto con sus debidas interpretaciones**.

1. Identifique y justifique el tipo y escala de medición (Clase03) de cada una de las variables contenidas en su base de datos y realice una descripción inicial de las variables cuantitativas mediante resúmenes numéricos (Clase05) e interprete los hallazgos que considere más relevantes. **(1 Punto)**
2. ¿En qué País se encuentra la mayor proporción de personas que poseen coronavirus y por qué podría deberse esto? **(1 Punto)**
3. ¿Podría pensarse que hay alguna relación entre tener Enfermedades Respiratorias y el Estado de Salud de los infectados por coronavirus? **(1 Punto)**
4. Cree una variable Índice de Masa Corporal, de la forma , y con ésta verifique si ¿Hay evidencia suficiente, que permita pensar que hay diferencias significativas entre el IMC de hombres y mujeres que poseen coronavirus? **(1 Punto)**
5. Presente el comportamiento de la curva de distribución de la Edad de los infectados, y muestre si ¿Realmente podría pensarse que la Edad influye a en el Estado de Salud de los infectados? **(1 Punto)**