**Estadística y Software R**

**Actividad Nº1**

**Ejercicio 1. Factores de riesgo de bajo peso al nacer**

El bajo peso al nacer, definido por un peso al nacer inferior a 2500 gr., ha sido una preocupación de los médicos durante años debido a que tanto las tasas de mortalidad como la de nacimientos defectuosos son muy altas para los niños con este problema. El comportamiento de la mujer durante el embarazo (incluyendo la dieta, los hábitos tabáquicos y los cuidados prenatales) pueden alterar las chances de un parto de un niño con bajo peso.

Los datos que se presentan en este ejercicio corresponden a 189 nacimientos de los cuales 59 han resultado en niños con bajo peso. El objetivo de este ejercicio es determinar cuáles de las variables presentes en la tabla de datos que se adjunta son factores de riesgo de bajo peso al nacer.

La tabla de datos que se presenta (archivo LOWBWT.xlsx) contiene las siguientes variables:

ID: Código de identificación

LOW: Bajo peso al nacer. (0 = ≥2500 g; 1 = <2500 g) (variable dependiente)

AGE: Edad de la madre

LWT: Peso de la madre el momento de la última menstruación (en libras)

RACE: Raza (1 = White; 2 = Black; 3 = Other)

SMOKE: Fumó durante el embarazo (0 = No 1 = Yes)

PTL: Antecedentes de embarazos prematuros (0 = None; 1 = One; 2 = Two, etc).

HT: Antecedentes de hipertensión arterial (0 = No; 1 = Yes)

UI: Irritabilidad uterina (0 = No; 1 = Yes)

FTV: Cantidad de consultas obstétricas durante el primer trimestre (0 = None; 1 = One; 2 = Two, etc.)

BWT: Peso al nacer del bebé en gramos

Se utilizará la información de este archivo para comenzar a utilizar el software R a través de RStudio.

**Primera parte**

1. Determinar cuáles de las variables de la tabla de datos son cuantitativas
2. Determinar cuáles de las variables de la tabla de datos son cualitativas

**Segunda parte. Creación de un Proyecto dentro de RStudio para este trabajo**

1. Crear una carpeta llamada “Curso\_r” en su directorio
2. Guardar el archivo “LOWBWT.xlsx” y el archivo “bajopeso.R” en dicha carpeta
3. Abrir RStudio
4. Crear un nuevo Proyecto asociado a la carpeta Curso\_r

**Tercera parte. Inicio de una sesión de trabajo dentro del Proyecto recientemente creado**

1. Importar a R el archivo LOWBWT.xlsx creando el data.frame “LOWBWT”
2. Visualizar el objeto “LOWBWT” (función View())
3. Cuantos casos tiene LOWBWT?
4. Cuantas variables tiene LOWBWT?
5. Qué variable ocupa la posición 10 en la tabla de datos?
6. Obtener una lista de los variables de LOWBWT (función names())
7. Hacer un resumen de las variables de LOWBWT (función summary())

**Cuarta parte. Manipular variables y casos en un data.frame, de acuerdo a las necesidades del trabajo.**

1. Crear un objeto data.frame que contenga sólo las variables ID, LOW y RACE.
2. Crear un objeto data.frame sólo con las pacientes que fumaron durante el embarazo.
3. Crear una variable con el peso de la madre en kg. (1 libra= 0.453592 kg.)
4. Crear una variable que considere nacimiento de bajo peso cuando el peso es menor de 2 kg. (variable dicotómica, 0 si peso es mayor o igual a 2 kg, 1 si peso es menor que 2 kg)
5. Crear una variable factor con la variable SMOKE
6. Crear una variable “**razanegra**” a partir de la variable RACE. **razanegra** debe valer uno si se es de raza negra y cero en cualquier otro caso.
7. Realizar todas las acciones pedidas en este apartado en una única instrucción
8. **Quinta parte. Aplicación de algunas funciones a las variables del archivo**
9. Calcular la media de la variable AGE y la variable LWT
10. Calcular utilizando la función summary() un resumen estadístico de la variable AGE y de la variable LWT

**EJERCICIO 2. DATOS CENSALES DE LOS DEPARTAMENTOS DE LA PROVINCIA DE CORDOA**

El archivo CORDOBA.SAV contiene información de algunas características de vivienda, hogar y población de departamentos de la provincia de Córdoba.

Este archivo consta de las siguientes variables:

* **DEPART**: Departamento provincial
* **NBI1**: Porcentaje de hogares con Necesidades Básicas Insatisfechas sobre el total de hogares de cada departamento.
* **NBI2**: Porcentaje de población en hogares con Necesidades Básicas Insatisfechas sobre el total de población de cada departamento.
* **CALMAT1**: porcentaje de viviendas que presentan materiales resistentes y sólidos en **todos** los paramentos (pisos, paredes o techos) e incorpora **todos** los elementos de aislación y terminación.
* **CALMAT2**: porcentaje de viviendas que presentan materiales resistentes y sólidos en **todos** los paramentos pero le faltan elementos de aislación o terminación **al menos en uno** de sus componentes (pisos, paredes, techos).
* **CALMAT3**: porcentaje de viviendas que presentan materiales resistentes y sólidos en **todos** los paramentos pero le faltan elementos de aislación o terminación en **todos** sus componentes, o bien presenta techos de chapa de metal o fibrocemento.
* **CALMAT4**: porcentaje de viviendas que presentan materiales no resistentes ni sólidos o de desecho **al menos en uno** de los paramentos.
* **ALFAB** : Porcentaje de alfabetos
* **COB** : Porcentaje de personas que tienen Obra social y/o plan de salud privado o mutual

Nota:

Las Necesidades Básicas Insatisfechas fueron definidas según la metodología utilizada en "La pobreza en la Argentina"

(Serie Estudios INDEC. N° 1, Buenos Aires, 1984).

Los hogares con Necesidades Básicas Insatisfechas (NBI) son los hogares que presentan al menos uno de los siguientes

indicadores de privación:

1- Hacinamiento: hogares que tuvieran más de tres personas por cuarto.

2- Vivienda: hogares en una vivienda de tipo inconveniente (pieza de inquilinato, vivienda precaria u otro tipo, lo que excluye

casa, departamento y rancho).

3- Condiciones sanitarias: hogares que no tuvieran ningún tipo de retrete.

4- Asistencia escolar: hogares que tuvieran algún niño en edad escolar (6 a 12 años) que no asistiera a la escuela.

5- Capacidad de subsistencia: hogares que tuvieran cuatro o más personas por miembro ocupado y, además, cuyo jefe no haya completado tercer grado de escolaridad primaria.

**CONSIGNA**

1. Determinar cuáles de las variables de la tabla de datos son cuantitativas
2. Determinar cuáles de las variables de la tabla de datos son cualitativas
3. Guardar el archivo “cordoba.sav”
4. Importar a R el archivo cordoba.sav creando el data.frame “cordoba”
5. Visualizar el objeto “cordoba” (función View())
6. Cuantos casos tiene cordoba?
7. Cuantas variables tiene cordoba?
8. Qué variable ocupa la posición 5 en la tabla de datos?
9. Obtener una lista de los variables de cordoba (función names())
10. Hacer un resumen de las variables de cordoba (función summary())
11. Calcular la media de la variable NBI1 y la variable NBI2
12. Calcular utilizando la función summary() un resumen estadístico de las variables de cordoba

**Ejercicio 3. Pacientes de ACV**

Los datos que se presentan en este ejercicio corresponden a 195 pacientes ingresados en un hospital de la zona.

La tabla de datos que se presenta (archivo ejercicio.xlsx) contiene las siguientes variables:

GUIA: Código de identificación

Edad: Edad del paciente

Sexo: Sexo del paciente

Vivo: Si sobrevivió al motivo de su internación

TAs: Tensión arterial sistólica

TAd: Tensión arterial diastólica

ACV: si el paciente ingresó con diagnóstico de ACV

ACVTIPO: Tipo de ACV, si corresponde

Se utilizará la información de este archivo para comenzar a utilizar el software R a través de RStudio.

1. Determinar cuáles de las variables de la tabla de datos son cuantitativas
2. Determinar cuáles de las variables de la tabla de datos son cualitativas
3. Guardar el archivo “ejercicio.xlsx” en la carpeta Curso\_r
4. Importar a R el archivo ejercicio.xlsx creando el data.frame “ejercicio”
5. Visualizar el objeto “ejercicio” (función View())
6. Cuantos datos tiene ejercicio?
7. Cuantas variables tiene ejercicio?
8. Qué variable ocupa la posición 2 en la tabla de datos?
9. Obtener una lista de los variables de ejercicio (función names())
10. Hacer un resumen de las variables de ejercicio (función summary())
11. Crear un objeto data.frame que contenga solo las variables GUIA, Edad y Sexo.
12. Crear un objeto data.frame mayores de 70 años.
13. Crear una variable con la Tensión Arterial media (TAm)
14. (TA media = TA diastólica + ((TA sistólica - TA diastólica) / 3)
15. Crear una variable que indique si el paciente tiene 70 o más años (variable dicotómica, 0 si edad es menor que 70 años, 1 si es mayor)
16. Crear una variable factor con la variable ACV.
17. Calcular la media de la variable Edad y las variables TAd y TAs
18. Calcular utilizando la función summary() un resumen estadístico de las mismas variables del punto anterior.