**PRUEBA 1 OCE-386**

**ANÁLISIS DE DATOS CON R**

**NOMBRE ALUMNOS:**

**FECHA ENTREGA:**

**NOMBRE PROFESOR EVALUADOR:**

**FECHA EVALUACIÓN:**

**Puntaje obtenido y nota**

**Puntaje total:**

**Nota:**

**Puntaje según nivel de logro.**

|  |  |
| --- | --- |
| **Nivel de logro** | **Puntaje** |
| Excelente | 4 |
| Muy bueno | 3 |
| Bueno | 2 |
| Suficiente | 1 |
| Insuficiente | 0 |

**Parámetros para el cálculo de la nota.**

|  |  |
| --- | --- |
| Nota máxima | 7,0 |
| Nota mínima | 1,0 |
| Nota aprobación | 4,0 |
| Exigencia | 60% |
| Puntaje máximo | 20 |

**INSTRUCCIONES GENERALES**

**1)** Elabore un **script .Rmd** usando el programa R-studio.cloud y guárdelo con el siguiente nombre: **PRUEBA\_1\_OCE-386\_dat\_1 , 2 o 3.Rmd según corresponda a los datos asignados para su tarea.**

2) Realice un análisis de datos con el set de datos asignado para la prueba.

3) Usando la librería knitr compile y genere un reporte en formato pdf. Envíe el reporte y el link de Rstudio.cloud por medio de la plataforma aula virtual.

4) Plazo máximo de entrega: viernes 10 de septiembre a las 23:59 PM.

**INSTRUCCIONES ESPECÍFICAS DEL CONTENIDO DEL REPORTE**

**Ejercicio 1. 4 puntos.**

Cree un objeto llamados **datos** e importe en el los datos asignados para la prueba usando la función read\_excel() de la librería “readxl”

Si la librería no está instalada proceda a su instalación siguiendo la guía en esta página web

<https://readxl.tidyverse.org/>

Identifique y clasifique todas las variables como continuas, discretas, ordinales, nominales o factores.

**Ejercicio 2. 4 puntos.**

Realice un análisis de datos univariado usando las funciones: summary(), str() y class(). Si corresponde, transforme las variables a factores o números usando la función as.factor() o as.numeric. Identifique la media, moda y la varianza de sus variables cuando corresponda.

**Ejercicio 3. 4 puntos.**

Elabore una tabla de frecuencia de 2x2 con las variables Genotipe, Family, Sex y Year\_class según su set de datos y responda la siguiente pregunta ¿Existe un número “balanceado” o “equivalente” de muestras por tratamiento?.

Use la función table(A,B)

<https://www.statmethods.net/stats/frequencies.html>

**Ejercicio 4. 4 puntos.**

Elabore un histograma, una gráfica de función de densidad empírica y una gráfica de qqplot de su variable independiente o respuesta (marcada en color verde) y responda la siguiente pregunta.

¿Qué función de distribución describe de mejor forma su variable dependiente? Normal, Bernoulli, Binomial.

**Ejercicio 5. 4 puntos.**

Realice una gráfica boxplot de su variable independiente en función de las otras variables bajo estudio y responda ¿Es adecuada esta función para describir a la variable dependiente? ¿Existen valores atípicos?.

Use la función boxplot()