

# Evaluación EDA

Diplomado en Análisis de datos con R para la acuicultura | Pontificia Universidad Católica de Valparaíso

Dr. José Gallardo Matus y Dra. María Angélica Rueda

03 June 2022

## PUNTAJE TOTAL Y NOTA

Table 1: Puntaje obtenido y nota.

Nombre	Puntaje obtenido	Nota
YESSICA LISETTE ORTEGA ASENCIOS	23	4.9

## PARÁMETROS PARA REALIZAR CÁLCULO DE LA NOTA

Table 2: Parámetros para el cálculo de la nota.

Item	Valor
Nota máxima	7,0
Nota mínima	1,0
Nota aprobación	4,0
Exigencia para aprobar	60 %
Puntaje máximo	32

Table 3: Puntaje según nivel de logro alcanzado en cada pregunta o ejercicio.

Nivel	Puntaje
Excelente	4
Muy bueno	3
Bueno	2
Suficiente	1
Insuficiente	0

## DETALLE DE PUNTAJE OBTENIDO POR PREGUNTA

Table 4: Puntaje obtenido por pregunta.

Pregunta	Puntaje
1 - Evidencias uso de control de versiones en github	0
2 - Archivo README en github	4
3 - Describe variación usando histogramas	4
4 - Identifica si los datos están balanceados	2
5 - Establece relación entre variables cuantitativas y factores usando gráficas de correlación, boxplot, interacción o de tamaño de los efectos.	4
6 - Identifica si existen errores, datos faltantes o valores atípicos	2
7 - Resume los datos usando tablas y estadística descriptiva.	3
8 - Utiliza librerías readxl, dplyr, ggplot2 o similar según naturaleza de los datos	4

Table 5: Evaluación cualitativa y retroalimentación EDA.

Retroalimentación con fines de aprendizaje.
<p>Buen trabajo, falta identificar si los datos estan balanceados y si existen o no valores atípicos para la variable discreta, datos faltantes y posibles errores en la toma de datos. No hay evidencia de control de versiones en Github. No realizar tablas de frecuencia de variables cuantitativas continuas</p> <p><code>table(Biom_1Peso)</code>, <code>table(Biom_7Talla)</code>, <code>table(Data_1PromedioPeso)</code>, <code>table(Data_8Promedio_Peso)</code>. Se sugiere ademas hacer tablas de frecuencia con dos variables de clasificación a la vez ej:</p> <p><code>table(Biom_8Dosis, Biom_6N_Estanque)</code> para saber si el diseño es balanceado entre estanques. No es conveniente hacer histogramas o graficas de correlación con el promedio de las variables</p> <p><code>pairs.panels(Data_8[,4:5])</code>, deben hacerse con los datos brutos. Si se presenta una grafica de dos factores para mostrar los datos <code>ggplot(Biom_8, aes(x=Dosis, y=Talla, fill=N_Estanque)) +</code></p>

`geom_boxplot()+ labs(y="Talla")`, no es necesario hacer boxplot de las mismas variables pero de un factor, es redundante. Elimine mensajes y alertas del reporte con `message=FALSE`, `warning=FALSE`. Agrupe tablas relacionadas con paneles (graficas de densidad). |