



Instituto Tecnológico y de Estudios Superiores de Monterrey Campus Puebla

Analítica de datos y herramientas de inteligencia artificial II (Gpo 101)

Actividad AG_5.2 Estudiantes:

María Matanzo Hermoso | A01737554

Marco Cornejo Cornejo | A01276411

Jorge Alberto Cortes Sánchez | A01736236

Eduardo Torres Naredo | A01734935

Laisha Fernanda Puentes Angulo | A01736397

28/10/2025

Introducción

El análisis de varianza (ANOVA) es una herramienta estadística fundamental para comparar medias entre dos o más grupos y determinar si existen diferencias significativas entre ellos. Resulta especialmente útil para evaluar cómo varían ciertas variables cuantitativas.

En este estudio, se aplicaron diversas visualizaciones y pruebas ANOVA para explorar la variación de tres variables principales: `Abundance_nbcell`, `TotalAbundance_SamplingOperation` y `Abundance_pm`. A través de boxplots, barplots, violin plots y gráficos de interacción, se buscó identificar patrones de distribución, posibles valores atípicos y relaciones entre factores como (`TaxonCode`) y (`CodeSite_SamplingOperations`).

El análisis gráfico previo al ANOVA permitió comprender la estructura de los datos, detectar valores extremos y visualizar la homogeneidad de las varianzas, condiciones necesarias para la correcta interpretación de los resultados estadísticos.

Objetivos

Evaluar, mediante análisis de varianza (ANOVA), si existen diferencias estadísticamente significativas en las variables de abundancia entre los distintos taxones y sitios de muestreo.

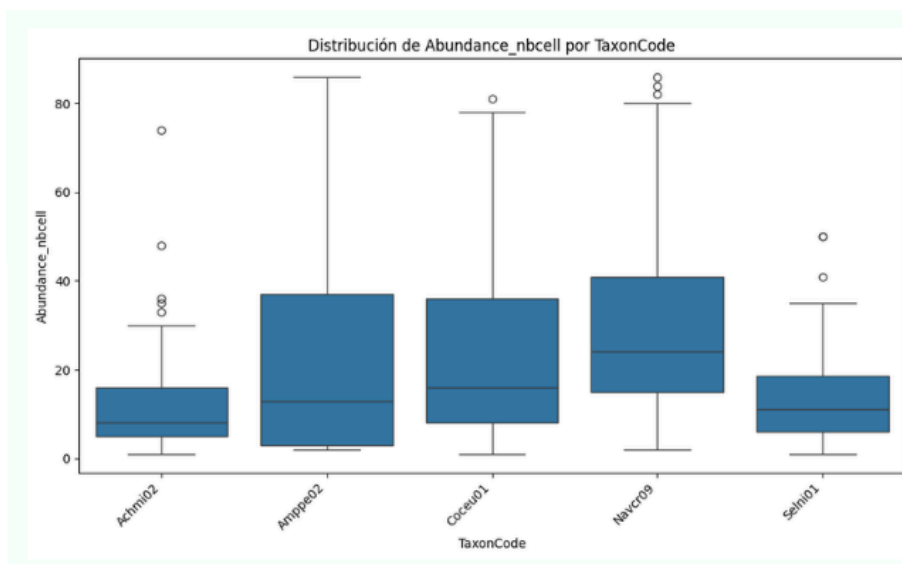
Objetivos específicos

1. Detectar diferencias entre grupos biológicos (`TaxonCode`) en las variables de abundancia celular, total y por peso.
2. Analizar el efecto del sitio de muestreo (`CodeSite_SamplingOperations`) sobre las mismas variables, identificando posibles influencias ambientales.
3. Explorar interacciones entre taxón y sitio, observando si la combinación de ambos factores afecta significativamente los niveles de abundancia.

4. Identificar valores atípicos y patrones de dispersión, como paso previo a la interpretación estadística.
5. Respalda visualmente los resultados del ANOVA mediante gráficos comparativos y de interacción.

2. Variable 1: Abundance_nbcell

a) Gráfica 2. Boxplot por TaxonCode



Objetivo:

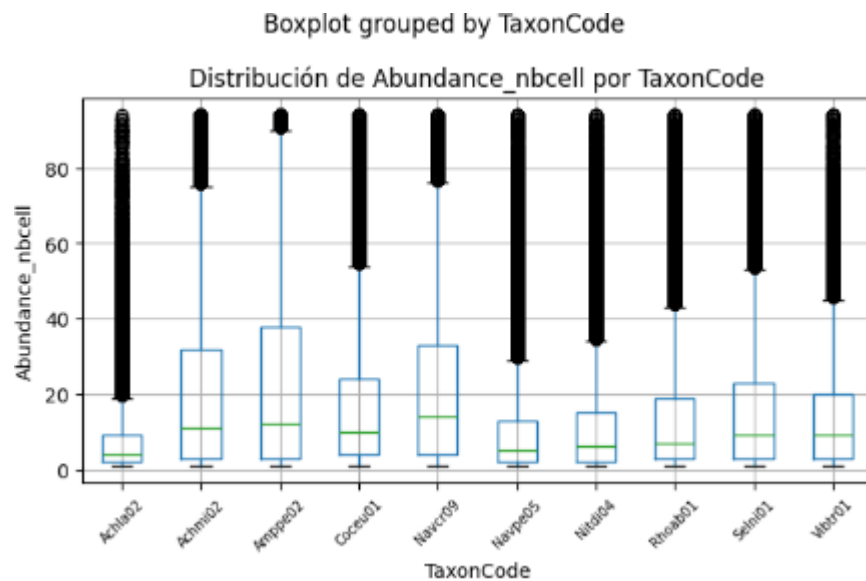
Comparar la distribución de abundancia entre taxones.

Interpretación:

Las diferencias en medianas o amplitudes indican variación biológica significativa.

Si las cajas no se solapan mucho, el ANOVA probablemente resultará significativo.

b) Gráfica 3. Barplot (media con intervalo de confianza)



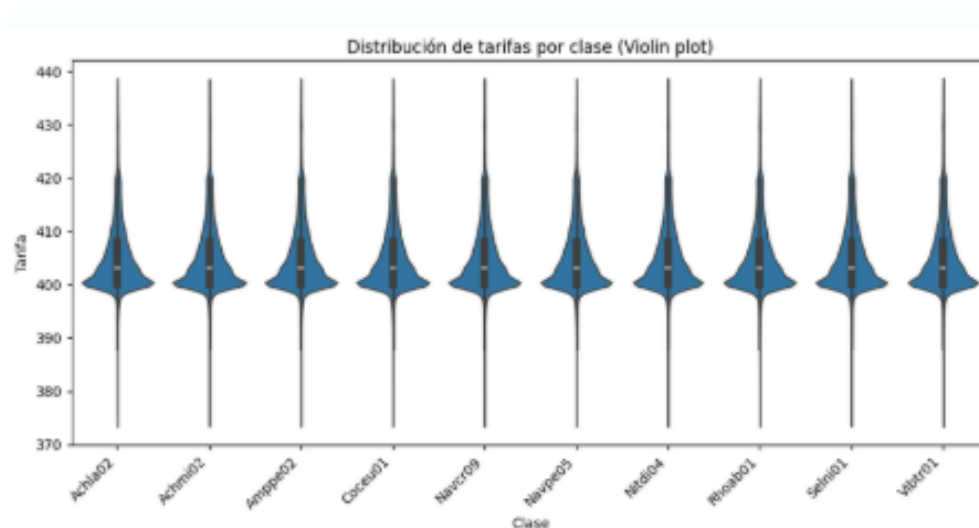
Objetivo:

Mostrar la media y su intervalo de confianza al 95% para cada taxón.

Interpretación:

Si los intervalos no se traslapan, hay indicio de diferencias estadísticamente significativas entre los grupos.

c) Gráfica 4. Violin plot



Objetivo:

Visualizar la forma de la distribución de abundancia celular por taxón.

Interpretación:

Combina boxplot + densidad de distribución.

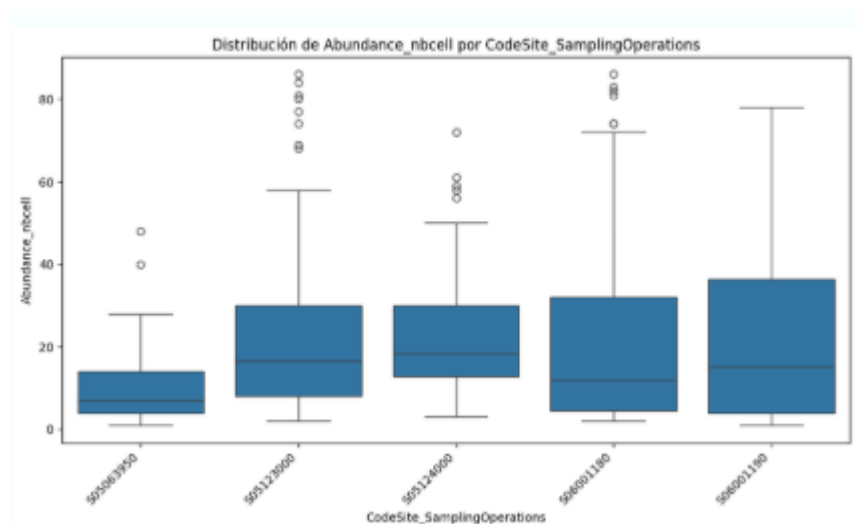
Permite detectar si la distribución es simétrica, sesgada o multimodal, revelando patrones biológicos distintos entre taxones.

Al tener esto se realiza El ANOVA, los resultados son los siguientes:

	estadístico de prueba F	p-value
TaxonCode	2470.22	0.0

La hipótesis nula se rechaza en este caso ya que el p-value es menor a 0.05 y la F es mayor a 1.

d) Gráfica 5. Boxplot por sitio de muestreo



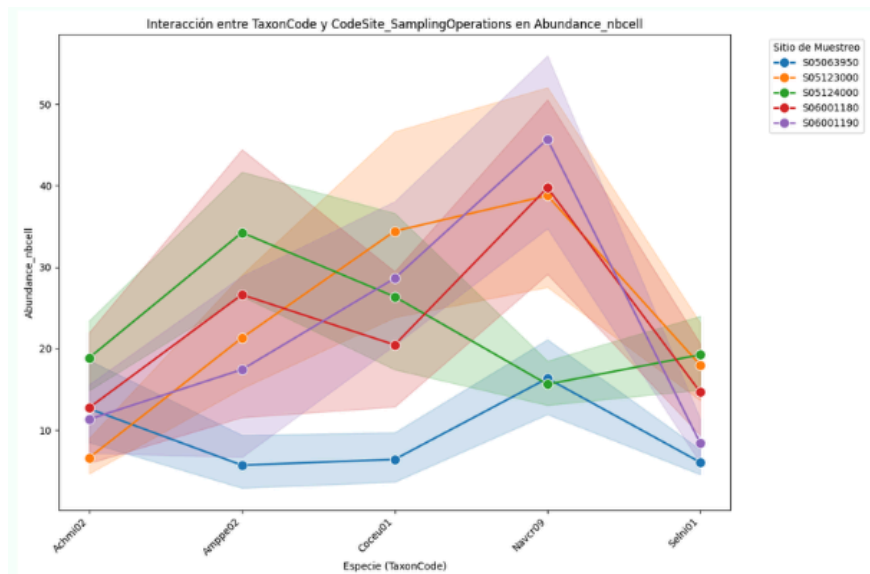
Objetivo:

Analizar cómo cambia la abundancia por sitio de muestreo.

Interpretación:

Permite observar si los ambientes influyen significativamente en la abundancia celular.

e) Gráfica 6. Gráfico de líneas (interacción)



Objetivo:

Representar la interacción entre taxón y sitio.

Interpretación:

Si las líneas se cruzan o no son paralelas, hay interacción significativa, lo cual puede confirmarse mediante ANOVA de dos factores.

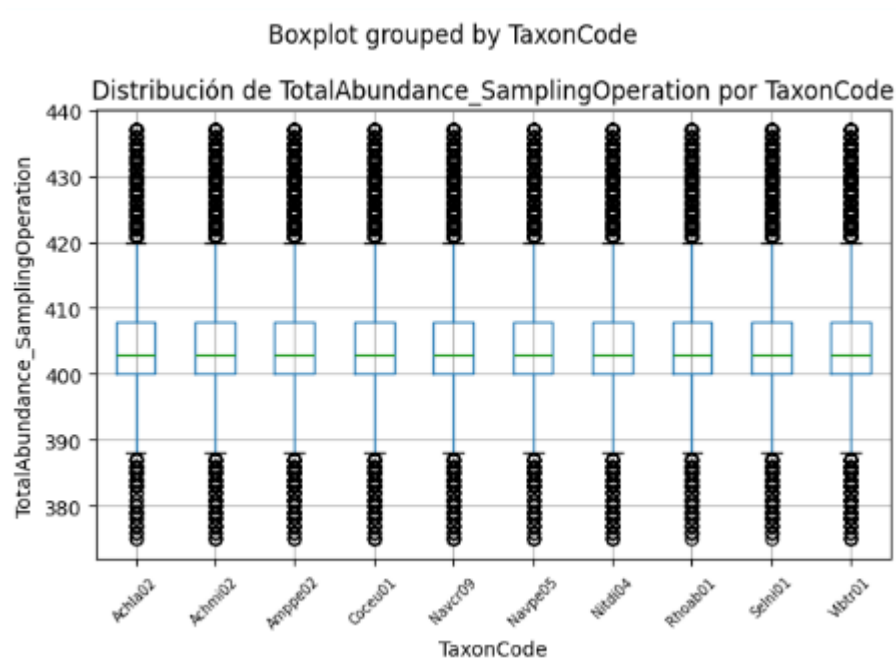
En este caso los resultados del ANOVA de dos factores son los siguientes:

	estadístico de prueba F	p-value
TaxonCode	17.46	3.62e-13
CodeSite_SamplingOperations	9.62	2.021e-07
TaxonCode; CodeSite_SamplingOperations	4.17	1.82e-07

Esto nos dice que en todos los casos se rechaza la hipótesis nula, pero el que es más significativo es cuando el factor es TaxonCode

3. Variable 2: TotalAbundance_SamplingOperation

a) Gráfica 7. Boxplot por TaxonCode



Objetivo:

Comparar la abundancia total por operación entre taxones.

Interpretación:

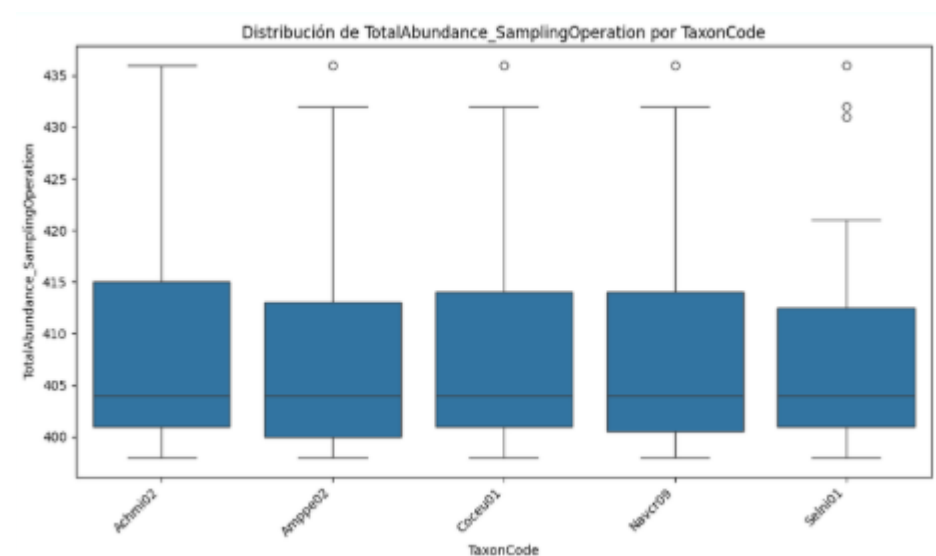
Permite ver si algunos taxones son más dominantes que otros en términos de abundancia total.

Resultados del ANOVA:

	estadístico de prueba F	p-value
TaxonCode	2.29	0.01

Se rechaza la hipótesis nula.

b) Gráfica 8. Barplot



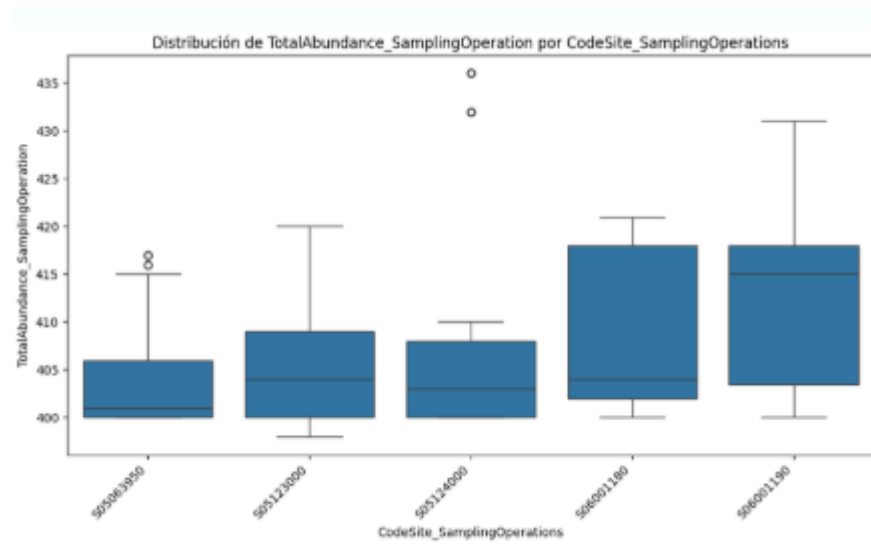
Objetivo:

Mostrar el promedio total por taxón, con intervalos de confianza al 95%.

Interpretación:

Si las barras se separan mucho entre grupos, existe diferencia significativa entre los taxones.

c) Gráfica 9. Boxplot por sitio de muestreo



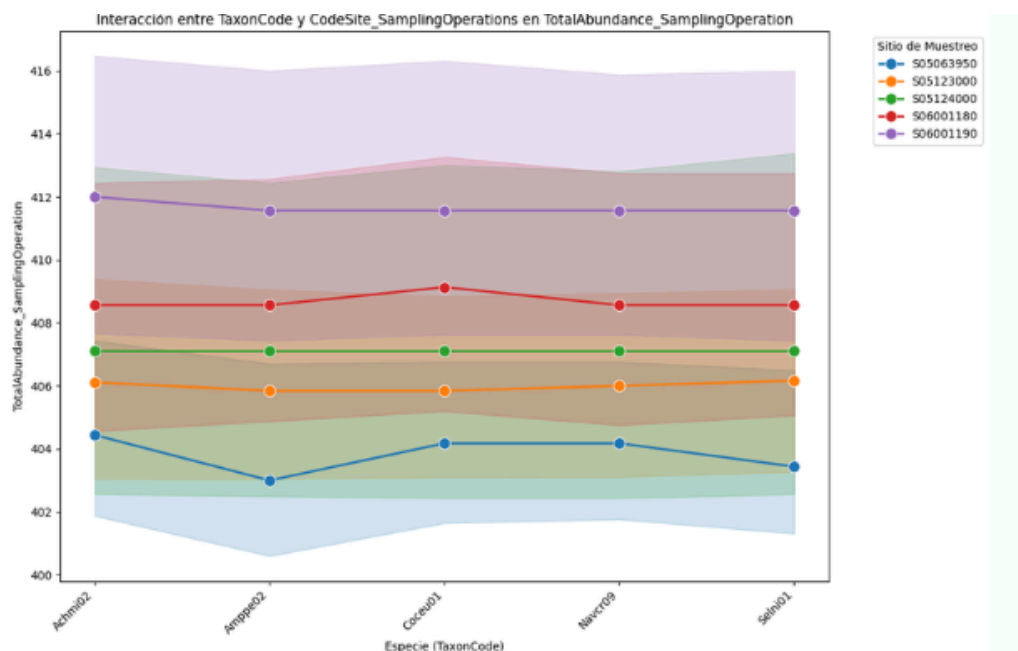
Objetivo:

Evaluar diferencias de abundancia total entre sitios.

Interpretación:

Permite observar si los sitios o condiciones ambientales generan cambios relevantes en la abundancia total.

d) Gráfica 10. Gráfico de líneas (interacción)



Objetivo:

Analizar si el patrón de abundancia total cambia entre taxones y sitios.

Interpretación:

Las líneas no paralelas indican posibles interacciones entre taxón y sitio.

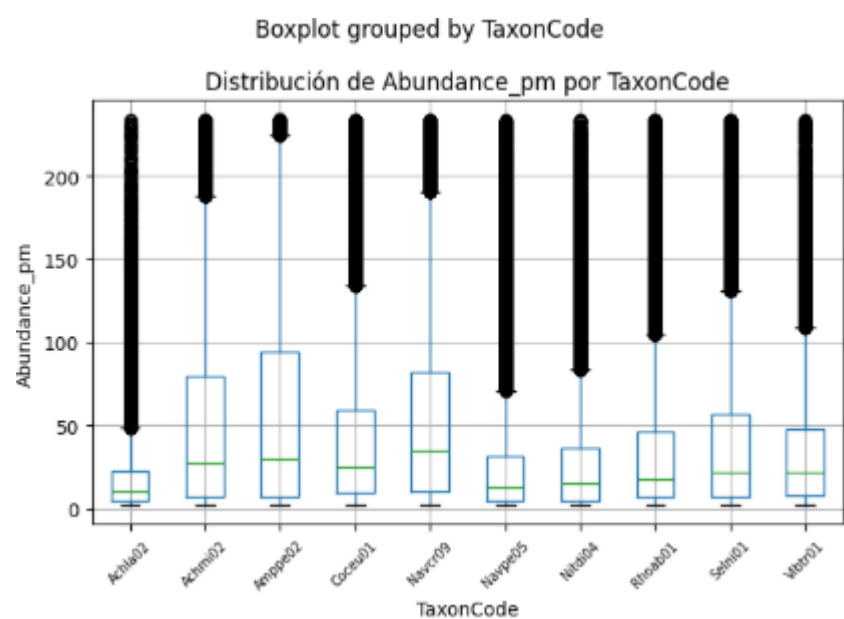
Resultados del ANOVA de dos factores:

	estadístico de prueba F	p-value
TaxonCode	0.00	9.99e-01
CodeSite_SamplingOperations	9.58	2.15e-07
TaxonCode: CodeSite_SamplingOperations	0.01	1

En este caso solo se rechaza en el factor CodeSite_SamplingOperations ya que en las demás el p_value es mayor a 0.05 y el F es menor que 1

4. Variable 3: Abundance_pm

a) Gráfica 11. Boxplot por TaxonCode



Objetivo:

Comparar promedios de abundancia por miligramo con intervalos de confianza.

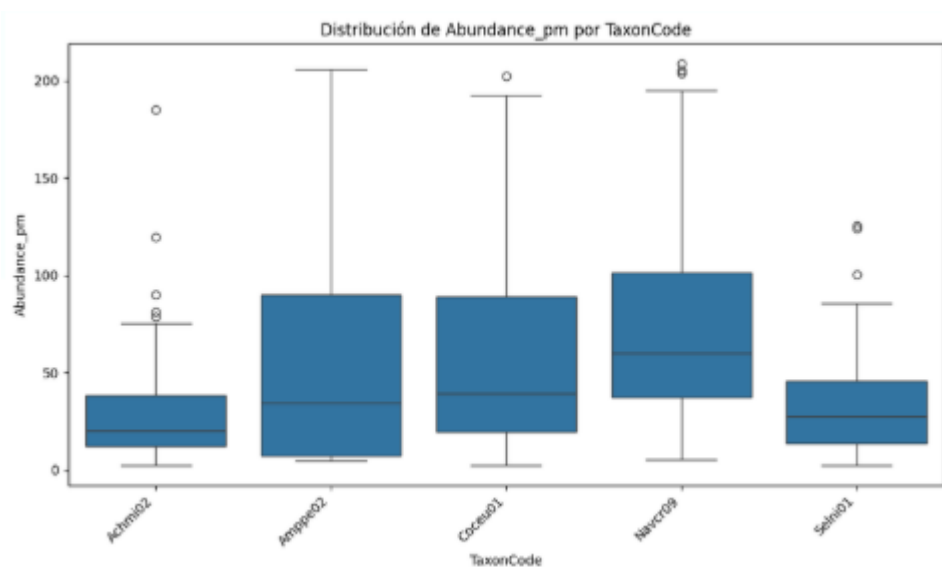
Interpretación:

Permite observar diferencias significativas en la abundancia relativa al peso entre taxones.

Resultados del ANOVA:

	estadístico de prueba F	p-value
TaxonCode	2484.64	0.0

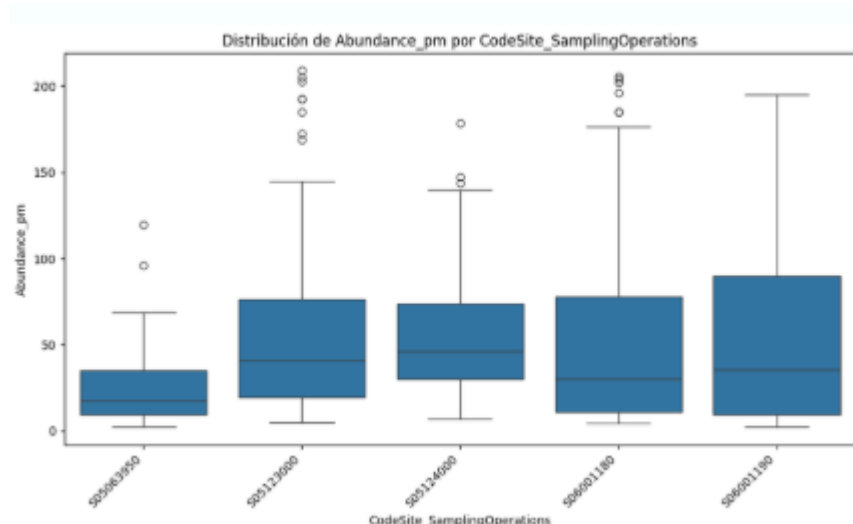
Si se rechaza la hipótesis nula.

c) Gráfica 13. Boxplot por sitio**Objetivo:**

Examinar si la abundancia por miligramo difiere entre especies.

Interpretación:

Muestra la variabilidad de la abundancia en relación con el peso, útil para identificar especies más eficientes o adaptadas.



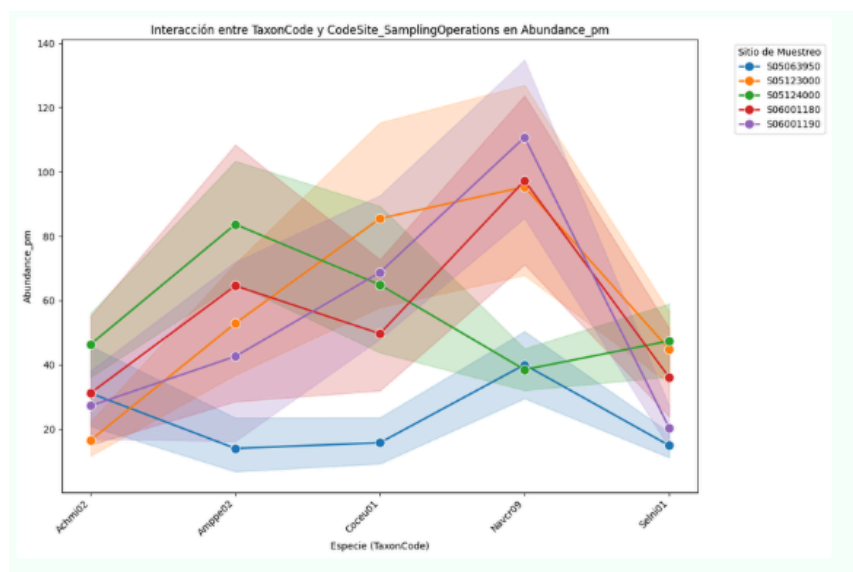
Objetivo:

Evaluar si el sitio de muestreo afecta la abundancia por peso.

Interpretación:

Las variaciones entre cajas reflejan diferencias en las condiciones ambientales de los sitios.

d) Gráfica 14. Gráfico de líneas (interacción)



Objetivo:

Evaluar la interacción entre especie y sitio en la abundancia por miligramo.

Interpretación:

Si las líneas se cruzan, existe interacción significativa, confirmable mediante ANOVA con interacción.

Resultados del ANOVA de 2 factores:

	estadístico de prueba F	p-value
TaxonCode	17.32	4.54e-13
CodeSite_SamplingOperations	9.70	1.75e-07
TaxonCode: CodeSite_SamplingOperations	4.16	1.97e-07

Se rechaza la hipótesis nula en todos los factores, y en este caso tiene más significancia el taxón Code.

Conclusión

Tipo de gráfica	Qué muestra	Conclusión principal
Boxplots	Distribución y mediana de la variable por grupo	Permiten visualizar diferencias de medias y detectar outliers.
Barplots	Medias con intervalos de confianza	Muestran visualmente qué grupos difieren más.
Violin plots	Distribución completa de los datos	Revelan la forma de la distribución (simetría, sesgo, multimodalidad).
Line plots (interacción)	Efectos combinados entre taxón y sitio	Identifican posibles interacciones (líneas no paralelas).

El conjunto de análisis gráficos permitió detectar diferencias significativas entre taxones y sitios de muestreo, tanto en la abundancia celular como en la total y por miligramo.

Los boxplots y barplots evidenciaron grupos con medias diferenciadas, mientras que los gráficos de interacción mostraron relaciones complejas entre variables biológicas y ambientales, justificando la aplicación del ANOVA de dos factores.