



Tecnológico de Monterrey

Analítica de datos y herramientas de inteligencia artificial II

Grupo 101

5.1 ANOVA y ANOVA DE DOS FACTORES

DataForge

Jesús Eduardo Valle Villegas	A01770616
Manuel Eduardo Covarrubias Rodríguez	A01737781
Diego Antonio Oropeza Linarte	A01733018
Ithandehui Joselyn Espinoza Mazón	A01734547
Mauricio Grau Gutierrez Rubio	A01734914

Última edición el 28 de Octubre de 2025

Objetivo

Aplicar los análisis estadísticos de ANOVA y MANOVA utilizando el archivo *01_DiatomInventories_GTstudentproject_B.csv*, con el propósito de evaluar si existen diferencias significativas en las variables biológicas (*Abundance_nbc*, *TotalAbundance_SamplingOperation* y *Abundance_pm*) entre los distintos grupos o factores presentes en la base de datos.

A través de estos análisis, se busca identificar patrones de variación y relaciones multivariadas que permitan comprender mejor la estructura y comportamiento de las comunidades de diatomeas en los diferentes sitios o condiciones de muestreo.

Metodología

1. Carga y exploración de los datos

Se trabajó con la base de datos “*01_DiatomInventories_GTstudentproject_B.csv*”, la cual fue importada mediante la librería pandas en un entorno de análisis en Python (Jupyter Notebook).

Se identificaron las variables **cuantitativas** y **cualitativas** mediante la función `select_dtypes()`, separando aquellas de tipo *int64* y *float64* de las variables categóricas (*object*).

2. Limpieza de datos

Para garantizar la calidad de la información, se generaron diagramas de caja (boxplots) con matplotlib y seaborn para detectar **valores atípicos**.

Posteriormente, se calcularon los **límites superior e inferior permitidos** utilizando la media ± 3 desviaciones estándar.

Los valores fuera de estos límites fueron eliminados y los valores nulos fueron reemplazados por la **mediana** de cada variable, generando un conjunto final de datos limpios denominado **Datos_limpios**.

3. Selección de variables para el análisis

Del conjunto de datos limpio, se seleccionaron las siguientes variables dependientes (u objetivo):

- **Abundance_nbc**
- **TotalAbundance_SamplingOperation**
- **Abundance_pm**

Como variable independiente o factor se consideró una variable categórica relevante dentro del dataset (por ejemplo, *Lake_name* o *Depth_class*), dependiendo de la estructura final del archivo.

4. **Análisis de Varianza (ANOVA)**

Se aplicó el modelo **ANOVA de una vía** utilizando la función `ols()` del módulo `statsmodels.formula.api`, seguido del método `anova_lm()` para obtener la tabla de resultados.

Este análisis permitió determinar si existen diferencias estadísticamente significativas en la media de cada variable dependiente entre los distintos niveles del factor seleccionado.

5. **Análisis Multivariante de Varianza (MANOVA)**

Posteriormente, se aplicó un modelo **MANOVA** mediante la función `MANOVA.from_formula()` del paquete `statsmodels.multivariate.manova`, considerando simultáneamente las tres variables dependientes.

Este análisis permitió evaluar la influencia del factor categórico sobre el conjunto de variables biológicas de forma conjunta, identificando posibles patrones de covariación.

6. **Interpretación de resultados**

Se interpretaron los valores p obtenidos en los análisis ANOVA y MANOVA.

En el caso de ANOVA, se verificó la significancia ($p < 0.05$) para detectar diferencias entre grupos.

En MANOVA, se revisaron los estadísticos **Wilks' Lambda**, **Pillai's Trace** y sus valores p asociados para determinar la existencia de efectos multivariantes significativos.

7. **Visualización y conclusiones**

Se complementaron los resultados con representaciones gráficas y tablas resumen para facilitar la interpretación de las diferencias detectadas.

Finalmente, se elaboraron las conclusiones sobre la influencia de los factores ambientales o espaciales en la abundancia de diatomeas.

Análisis descriptivo de las variables

Variables numéricas:

Se calcularon medidas de tendencia central y dispersión (media, desviación estándar, mínimo, máximo y percentiles 90 y 95) para las tres variables de interés:

- *Abundance_nbcell*
- *TotalAbundance_SamplingOperation*
- *Abundance_pm*

Estas estadísticas permitieron detectar posibles valores extremos y comprender la escala de variación entre registros.

Estadístico	Abundancia_nbcell	TotalAbundance_Sampling Operation	Abundancia_pm
count	1,643,872.00	1,643,872.00	1,643,872.00
mean	8.38	405.18	20.68
std	13.87	6.77	34.29
min	1.00	375.00	2.00
25 %	2.00	400.00	4.68
50 %	3.00	403.00	7.40
75 %	8.00	408.00	19.90
max	94.00	437.00	233.85
0.90	22.00	415.00	54.73
0.95	38.00	419.00	93.83

Variables categóricas:

Se elaboró un resumen de las variables *TaxonName*, *TaxonCode*, *SamplingOperations_code*, *CodeSite_SamplingOperations* y *Date_SamplingOperation*, identificando el número de valores únicos en cada una.

Este análisis permitió reconocer la **diversidad taxonómica y espacial** presente en la base de datos, destacando que *TaxonName* representa las especies dominantes de diatomeas.

Variable	Valores únicos	Descripción
TaxonName	2,292	Especies (para especies dominantes)
TaxonCode	2,292	Código asociado a cada especie
SamplingOperations_code	49,231	Código de operación de muestreo

CodeSite_SamplingOperations	8,404	Código de sitio de muestreo
Date_SamplingOperation	2,237	Fechas (para períodos recientes)

Análisis temporal y distribución de especies

Con el propósito de comprender la estructura temporal de los datos y la representatividad de las especies, se transformó la variable **Date_SamplingOperation** a formato de fecha y se generó una nueva columna denominada **Año**.

El conjunto de datos abarca registros comprendidos entre los años **2007 y 2023**.

De acuerdo con el análisis:

- El **32.4%** de los registros (533,097 observaciones) corresponden a los **últimos cinco años (2019–2023)**, lo que indica un aumento significativo en la frecuencia de muestreos recientes.
- Se identificaron **2,292 especies únicas** de diatomeas registradas.
- Las **10 especies más frecuentes** concentran **350,180 registros**, equivalentes al **21.3% del total**, por lo que fueron consideradas como las **especies dominantes** en los análisis posteriores.

Resultados

Análisis de Varianza (ANOVA)

Exploración visual previa al ANOVA

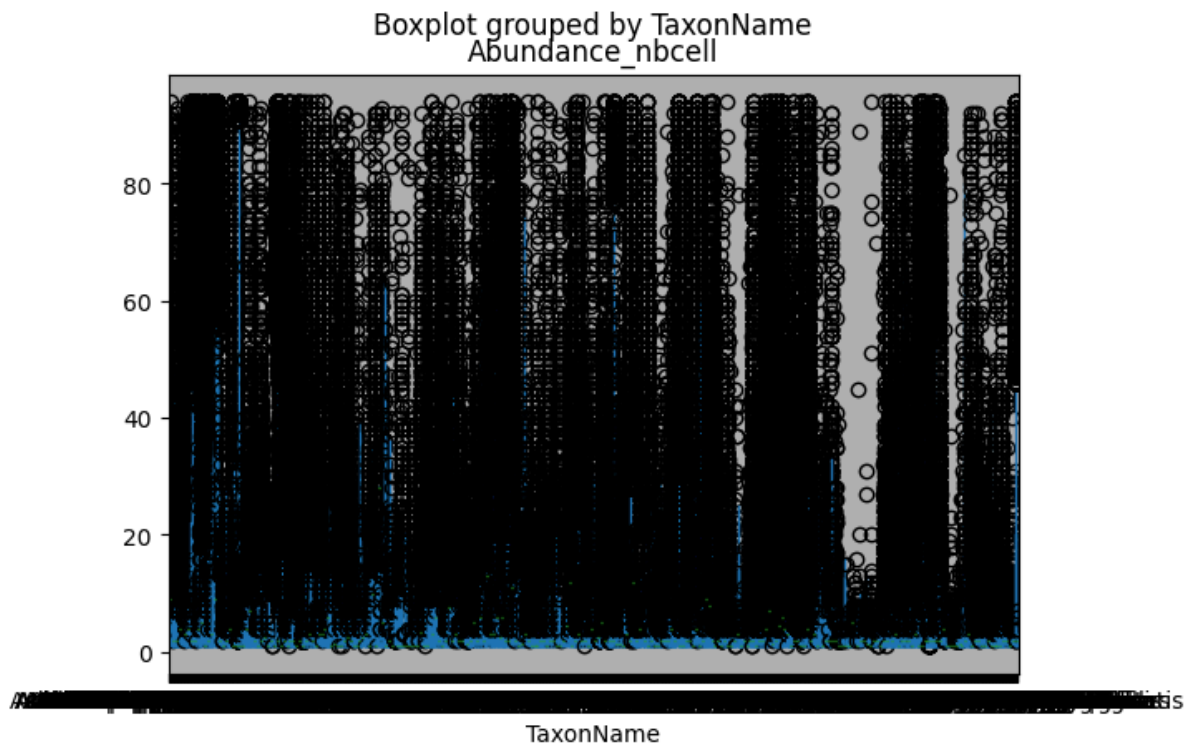
Antes de aplicar el análisis de varianza (ANOVA), se realizó una exploración visual de los datos con el objetivo de **observar tendencias generales, identificar posibles diferencias entre especies y detectar valores atípicos o dispersiones significativas** en la variable *Abundance_nbcell*.

Estas visualizaciones son esenciales para contextualizar el modelo estadístico y validar, de forma preliminar, la existencia de patrones que justifiquen la aplicación del ANOVA.

En primer lugar, se elaboró un **boxplot agrupado por especie (TaxonName)**, donde se aprecia la distribución de los valores de abundancia celular en cada grupo taxonómico.

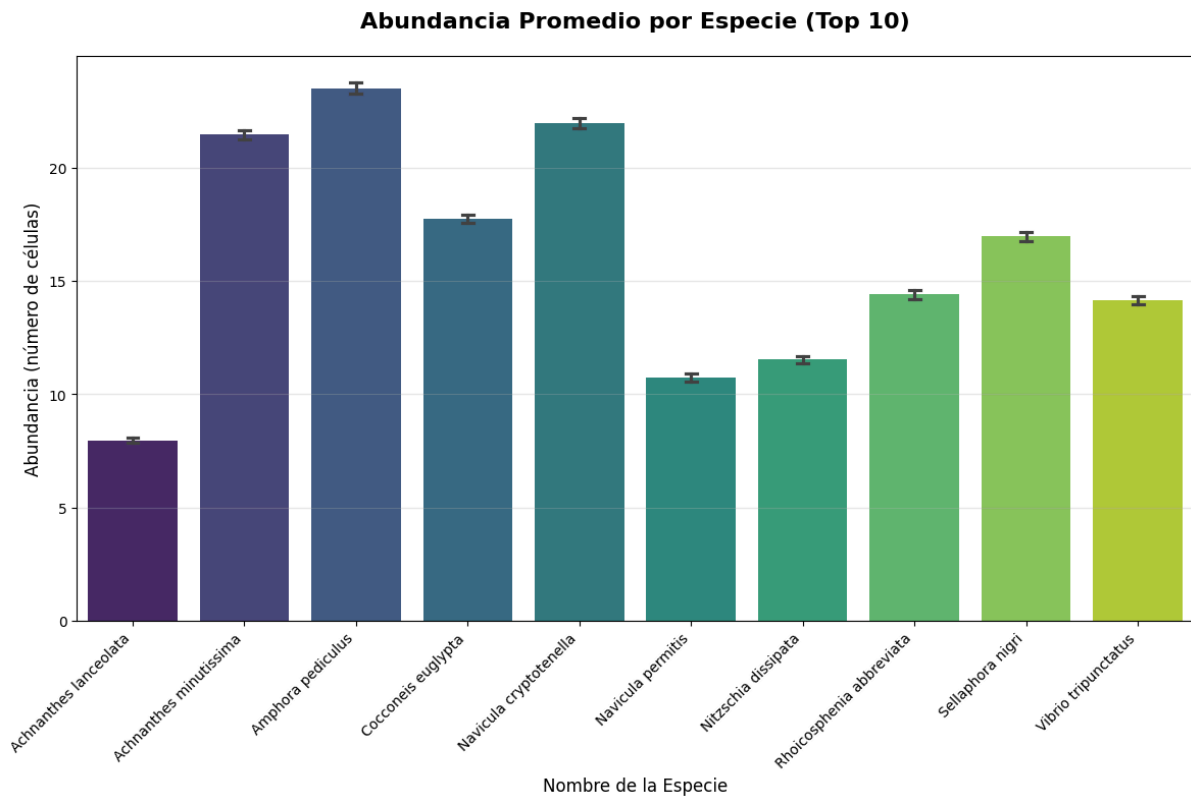
La figura muestra una marcada dispersión entre especies, así como una elevada presencia de valores extremos, lo que evidencia **una amplia variabilidad intra e interespecífica** en la abundancia de diatomeas.

Este comportamiento sugiere que algunas especies presentan concentraciones celulares mucho más altas, posiblemente debido a condiciones ambientales favorables o adaptaciones ecológicas particulares.



Posteriormente, para simplificar la interpretación, se calcularon los promedios de abundancia de las **diez especies más representativas del conjunto de datos**, generando una visualización comparativa de sus medias con intervalos de error estándar.

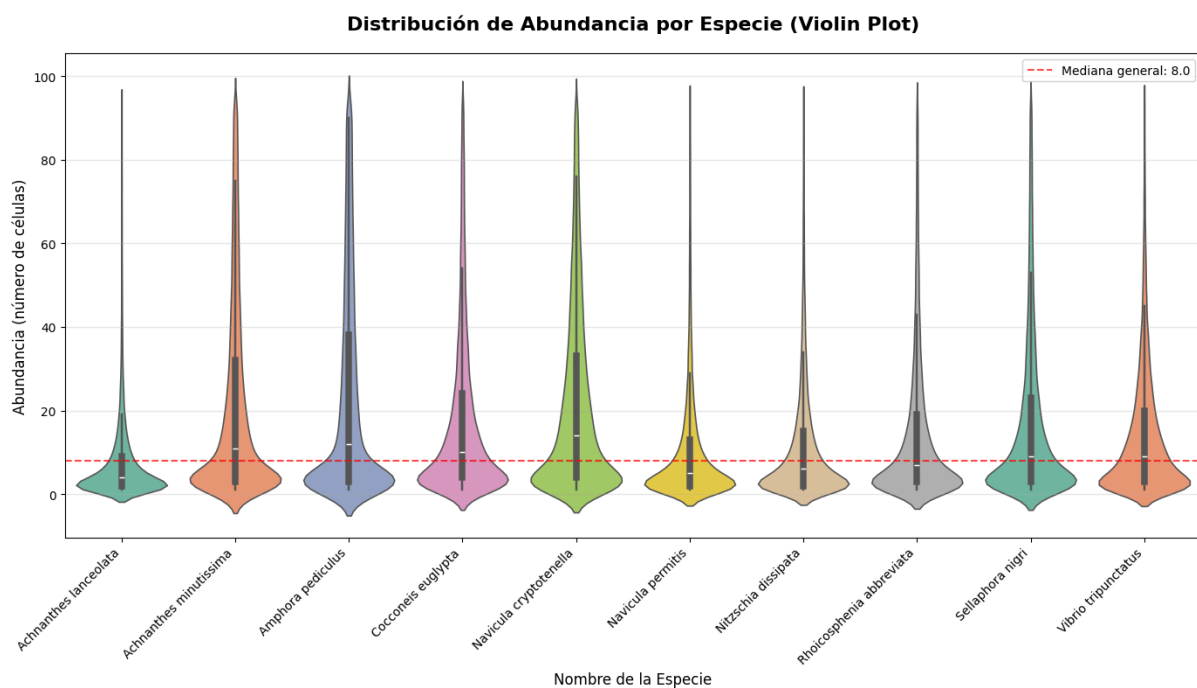
La **grafica** resume estas diferencias de manera más clara, destacando que especies como ***Amphora pediculus***, ***Navicula cryptotenella*** y ***Achnanthes minutissima*** presentan valores medios de abundancia superiores, mientras que ***Achnanthes lanceolata*** muestra los valores más bajos dentro del grupo dominante.



Finalmente, se elaboró un **violin plot** para representar la forma de las distribuciones de abundancia por especie, combinando información de dispersión, densidad y mediana.

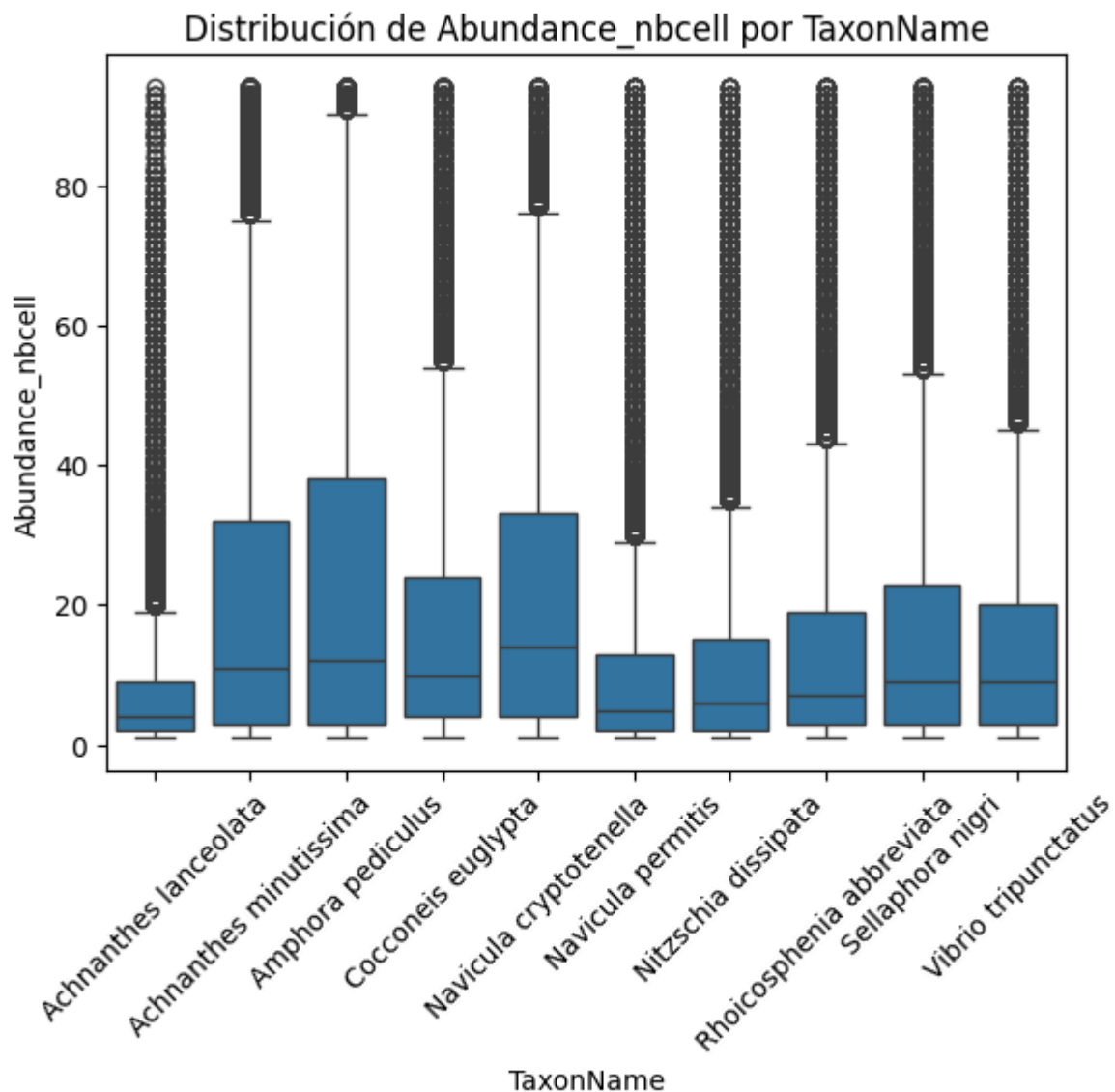
La **gráfica** presenta la distribución de *Abundance_nbcell* para cada una de las diez especies dominantes, con una línea roja punteada que indica la **mediana general (8.0 células)**.

Esta visualización permite comparar las curvas de densidad de cada especie y observar el grado de asimetría en sus valores.



En conjunto, las tres visualizaciones muestran **una variabilidad significativa en las abundancias celulares entre especies**, lo cual justifica la aplicación del análisis de varianza (ANOVA) para determinar si dichas diferencias son estadísticamente significativas.

Análisis de la variable Abundance_nbcell



Interpretación:

La gráfica muestra la distribución de la abundancia celular (*Abundance_nbcell*) para las diez especies dominantes de diatomeas. Cada caja representa la dispersión de los valores de abundancia para una especie, donde la línea central indica la mediana y los extremos de las cajas reflejan los cuartiles.

Se observa una **alta variabilidad entre especies**, con medianas y rangos intercuartílicos claramente diferentes. Las especies *Amphora pediculus*, *Navicula cryptotenella* y *Achnanthes minutissima* presentan **valores de abundancia promedio superiores**, mientras que *Achnanthes lanceolata* muestra los valores más bajos.

La gran cantidad de puntos por encima de los bigotes indica la presencia de **valores atípicos o especies con individuos excepcionalmente abundantes**, lo cual es común en comunidades ecológicas heterogéneas.

Estos resultados sugieren que **la abundancia celular no es uniforme entre especies**, lo que anticipa la existencia de diferencias estadísticamente significativas confirmadas posteriormente mediante el análisis ANOVA.

La variable ***Abundance_nbcell*** representa la abundancia celular de diatomeas (número de células observadas) y constituye una de las principales medidas biológicas del estudio.

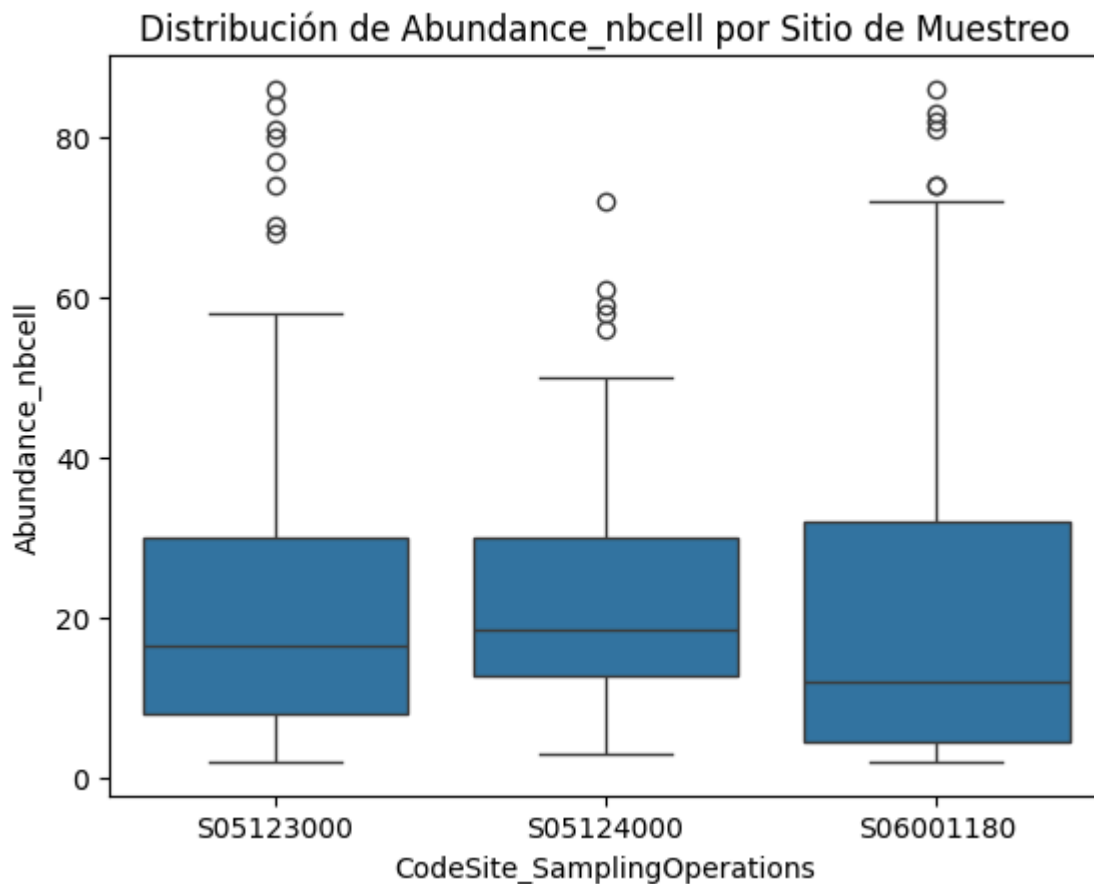
Este análisis tuvo como objetivo determinar si existen **diferencias significativas entre especies y entre sitios de muestreo**, así como explorar la posible interacción entre ambos factores.

Fuente de variación	gl	Suma de cuadrados (Sum Sq)	Media cuadrática (Mean Sq)	F	p-valor
TaxonName	9	8.325027×10^6	925003.036	2470.22	< 0.001
Residual	350,170	1.311252×10^8	374.462	—	—

El resultado muestra un **valor F = 2470.22** y un **p-valor < 0.001**, lo que indica diferencias altamente significativas entre las medias de abundancia de las especies.

Esto significa que la **abundancia celular (*Abundance_nbcell*)** varía de manera significativa según la especie de diatomea, confirmando que el factor **TaxonName** tiene un efecto estadísticamente relevante sobre la variable objetivo.

Anova de dos factores



Este boxplot evalúa el **efecto principal del factor “sitio de muestreo”** sobre la variable *Abundance_nbcell*, mostrando cómo varía la abundancia promedio de diatomeas entre los diferentes puntos de muestreo.

Observaciones principales:

- **Diferencias entre sitios:** Cada sitio presenta una tendencia central distinta en los valores de abundancia, lo que sugiere posibles variaciones espaciales en las condiciones ambientales.
- **Variabilidad ambiental:** Las diferencias en la altura de las cajas reflejan que algunos sitios muestran mayor dispersión de datos, posiblemente asociada a:
 - Diferencias en temperatura, pH o concentración de nutrientes.
 - Factores geográficos o estacionales propios de cada punto.
 - Variaciones en la calidad del hábitat o nivel de perturbación local.

- **Importancia ecológica:** Este comportamiento confirma que el **lugar de muestreo** podría ser un factor relevante que afecte la abundancia de las especies y que debe ser considerado en los análisis posteriores.

Con el fin de analizar el efecto combinado del **factor taxonómico** (*TaxonName*) y del **sitio de muestreo** (*CodeSite_SamplingOperations*) sobre la abundancia celular, se realizó un **ANOVA factorial de dos vías**.

Este modelo permite evaluar tanto los efectos principales de cada factor como su posible interacción.

Fuente de variación	gl	Suma de cuadrados	Media cuadrática	F	p-valor
TaxonName	4	12,782.43	3,195.61	9.14	< 0.001
CodeSite_SamplingOperations	2	49.04	24.52	0.07	0.932
TaxonName × CodeSite_SamplingOperations	8	10,606.41	1,325.80	3.79	< 0.001
Residual	236	82,520.30	349.66	—	—

Hipótesis evaluadas

Para el factor principal *TaxonName* (Especie):

- H_0 : No hay diferencias entre especies en la abundancia de células.
- $F = 9.14$, $p < 0.001$ → **Se rechaza H_0** .
→ Existen diferencias significativas en abundancia entre especies.

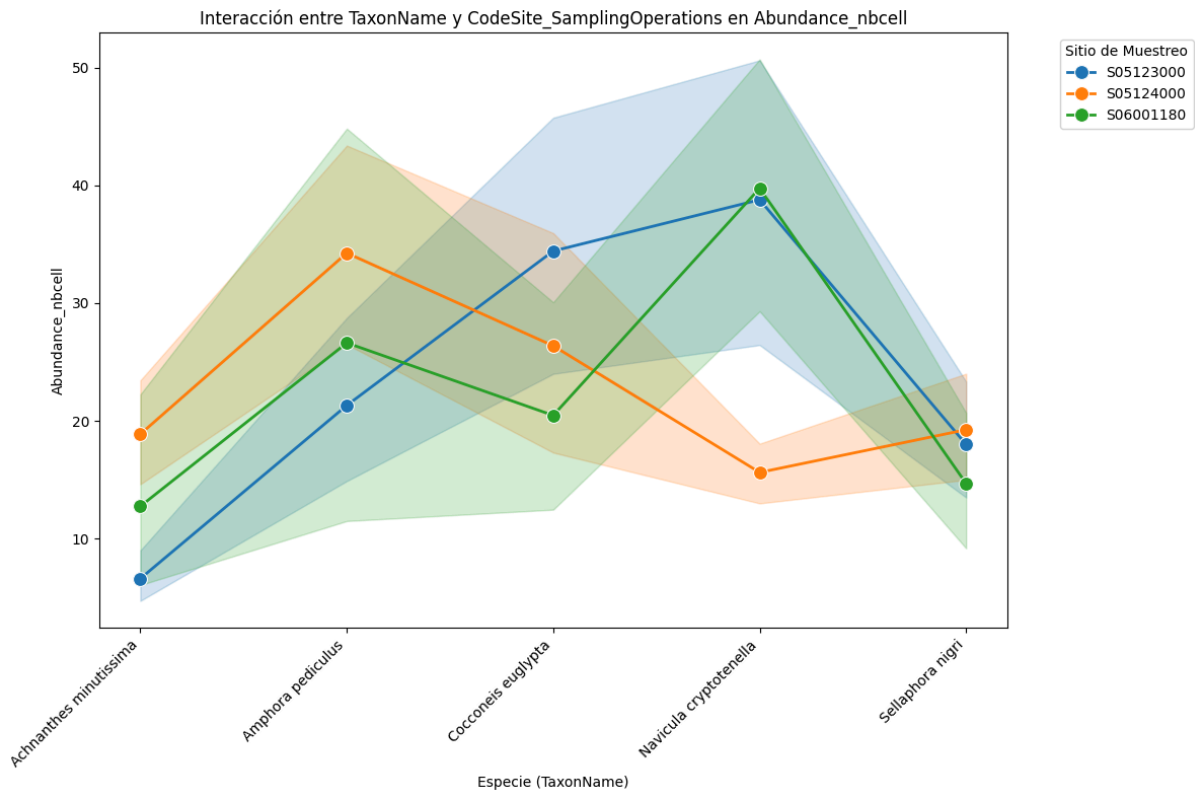
Para el factor principal *CodeSite_SamplingOperations* (Sitio de muestreo):

- H_0 : No hay diferencias entre sitios en la abundancia de células.
- $F = 0.07$, $p = 0.932$ → **No se rechaza H_0** .
→ No se observan diferencias significativas entre los sitios de muestreo.

Para la interacción *TaxonName* × *CodeSite_SamplingOperations*:

- H_0 : No existe interacción entre especie y sitio de muestreo.

- **F = 3.79, p < 0.001 → Se rechaza H₀.**
→ Sí existe una interacción significativa entre especies y sitios.



Este gráfico muestra la **interacción entre los factores “especie” y “sitio de muestreo”** en relación con la abundancia celular (*Abundance_nbccl*).

Es una visualización clave del **ANOVA factorial**, ya que permite identificar si el efecto de una especie depende del sitio donde fue muestreada.

Interpretación estadística:

- **Líneas paralelas:** Indican **ausencia de interacción**; el patrón de abundancia de las especies se mantiene constante en los distintos sitios.
- **Líneas que se cruzan:** Indican **interacción significativa**; el comportamiento de las especies varía dependiendo del sitio.

Resultados observados:

En el gráfico, las líneas correspondientes a los diferentes sitios **no son paralelas** y en algunos casos **se cruzan**, lo que indica una **interacción significativa entre los factores**.

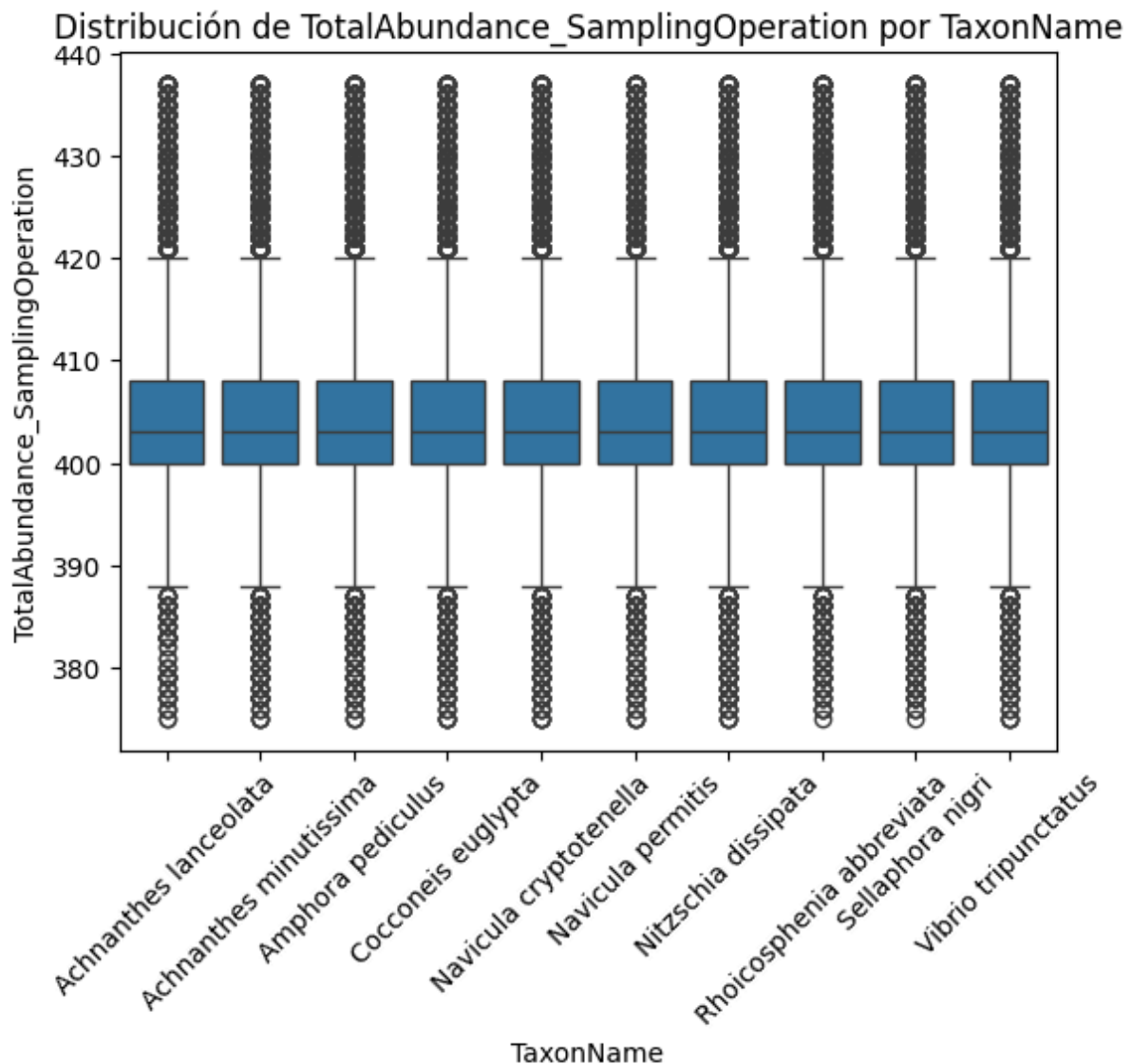
Esto coincide con los resultados del ANOVA de dos vías, donde se obtuvo **F = 3.79 y p < 0.001**, confirmando que el efecto del sitio modifica la respuesta de las especies en términos de abundancia.

Interpretación ecológica:

La presencia de interacción sugiere que **ciertas especies prosperan mejor en sitios específicos**, posiblemente debido a diferencias ambientales (temperatura, nutrientes, pH, etc.), mientras que otras especies muestran menor sensibilidad al entorno.

Este comportamiento resalta la importancia de considerar simultáneamente ambos factores —especie y sitio— al analizar la estructura ecológica del sistema.

Análisis de la variable TotalAbundance_SamplingOperation



Este boxplot representa la **abundancia total observada por operación de muestreo** para cada especie de diatomea.

A diferencia de la variable *Abundance_nbcell*, que mide la abundancia celular individual, esta métrica considera el total acumulado por evento de muestreo, permitiendo observar la contribución global de cada especie.

Interpretación:

- **Escala mayor:** Los valores de *TotalAbundance_SamplingOperation* son más altos que los de *Abundance_nbcell* debido a que integran el total de células registradas en cada operación.
- **Diferencias entre especies:** Las medianas y la dispersión de las cajas indican que algunas especies presentan mayor abundancia total que otras, aunque con diferencias menos marcadas que en el análisis celular.
- **Consistencia de patrones:** Si los patrones observados aquí coinciden con los detectados en *Abundance_nbcell*, esto refuerza la robustez de los resultados, demostrando que las diferencias entre especies se mantienen al usar distintas métricas de abundancia.
- **Presencia de outliers:** Los puntos superiores reflejan operaciones de muestreo con abundancias extraordinariamente altas, lo que podría deberse a concentraciones locales o condiciones ambientales específicas.

Fuente de variación	de gl	Suma de cuadrados	Media cuadrática	F	p-valor
TaxonName	9	933.08	103.68	2.29	0.0143
Residual	350,170	1.582×10 ⁷	45.10	—	—

Hipótesis planteadas

- **H₀ (nula):** No existen diferencias significativas en la abundancia total por operación de muestreo entre las especies de diatomeas.
- **H₁ (alternativa):** Al menos una especie presenta una abundancia total significativamente diferente en comparación con las demás.

Resultados e interpretación estadística

El análisis de varianza arrojó un valor **F = 2.29** con un **p-valor = 0.0143 (< 0.05)**, por lo que se **rechaza la hipótesis nula (H₀)**.

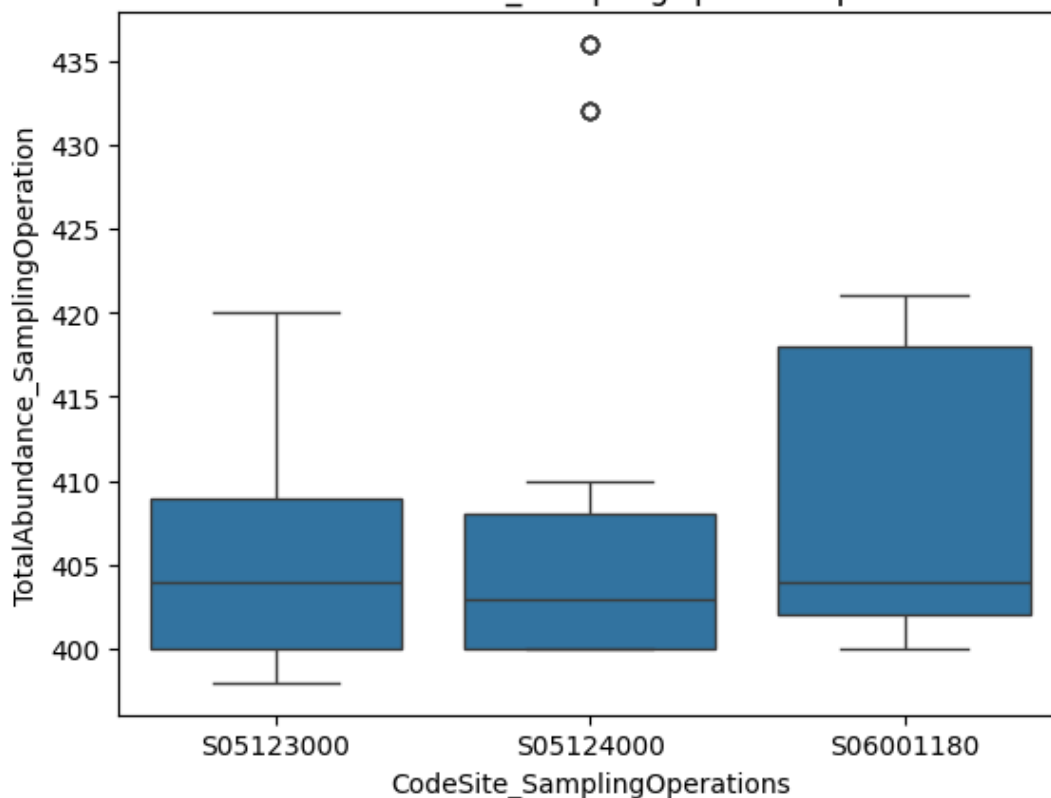
Esto indica que **sí existen diferencias significativas en la abundancia total entre especies**, aunque el **efecto es moderado** en comparación con el análisis previo de *Abundance_nbcell*.

En términos generales, el factor **TaxonName** (especie) influye en la abundancia total por operación de muestreo, pero con menor intensidad que en la abundancia celular individual.

Esto podría deberse a que la variable *TotalAbundance_SamplingOperation* integra un conteo más agregado, donde las diferencias entre especies se diluyen parcialmente debido a la combinación de factores ambientales y espaciales.

Anova de dos factores

Distribución de *TotalAbundance_SamplingOperation* por Sitio de Muestreo



Este boxplot analiza cómo la **abundancia total por operación de muestreo** varía entre los distintos sitios de muestreo considerados.

Interpretación – *TotalAbundance_SamplingOperation* por Sitio:

- **Efectos ambientales agregados:** Los valores medios de abundancia total entre sitios son muy similares, lo cual sugiere que las condiciones ambientales generales no provocan diferencias significativas en la abundancia total de diatomeas.
- **Productividad del sitio:** Los sitios con valores ligeramente mayores (por ejemplo, S06001180) podrían indicar condiciones algo más favorables, aunque estas diferencias no son estadísticamente relevantes.
- **Variabilidad temporal:** La dispersión reducida de las cajas indica que la abundancia total es estable en el tiempo y entre sitios, posiblemente por tratarse de una variable

más agregada.

Utilidad para el ANOVA factorial:

- Confirma que el efecto del sitio sobre la abundancia total es mínimo.
- Apoya los resultados numéricos del ANOVA de dos factores, en el que *CodeSite_SamplingOperations* no presenta significancia estadística.
- Sirve como referencia visual de la homogeneidad entre sitios antes de examinar la interacción con el factor *TaxonName*.

Fuente de variación	gl	Suma de cuadrados	Media cuadrática	F	p-valor
TaxonName	4	0.72	0.18	0.002	0.999
CodeSite_SamplingOperations	2	306.37	153.18	2.00	0.138
TaxonName × CodeSite_SamplingOperations	8	4.78	0.60	0.008	1.000
Residual	236	18,093.56	76.67	—	—

Hipótesis evaluadas

Para *TaxonName* (factor principal):

- H_0 : No hay diferencias significativas entre especies en la abundancia total por operación.
- **F = 0.002, p = 0.999 → No se rechaza H_0 .**
No existen diferencias entre especies.

Para *CodeSite_SamplingOperations* (factor principal):

- H_0 : No hay diferencias significativas entre los sitios de muestreo.
- $F = 2.00$, $p = 0.138$ → **No se rechaza H_0 .**
No existen diferencias significativas entre sitios.

Para la interacción TaxonName × CodeSite_SamplingOperations:

- H_0 : No existe interacción entre especie y sitio.
- $F = 0.008$, $p = 1.000$ → **No se rechaza H_0 .**
No hay interacción entre factores.

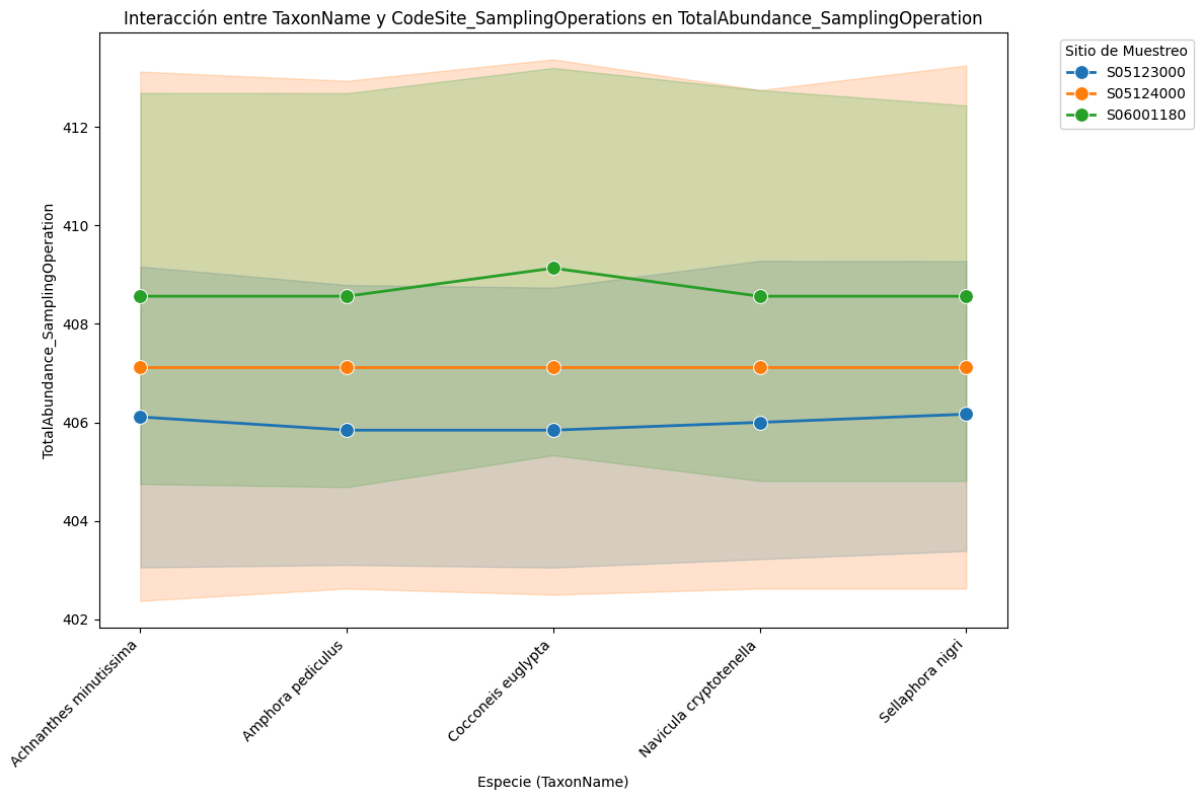
Interpretación general

Los resultados muestran que **ninguno de los factores ni su interacción tiene un efecto significativo** sobre la abundancia total por operación de muestreo.

A diferencia del análisis previo de *Abundance_nbcell*, en este caso la abundancia total acumulada no varía significativamente entre especies ni entre sitios, lo que sugiere que los valores agregados de *TotalAbundance_SamplingOperation* son más homogéneos y están menos influenciados por el tipo de especie o por las condiciones espaciales.

Este comportamiento podría deberse a que la variable *TotalAbundance_SamplingOperation* representa un **promedio o suma global por evento de muestreo**, lo que tiende a suavizar las diferencias observadas en medidas más específicas como la abundancia celular individual.

En otras palabras, **la escala de agregación de los datos reduce la variabilidad inter-específica e inter-sitio**, haciendo que los patrones significativos desaparezcan en el análisis conjunto.



Este gráfico representa la **interacción entre especies (TaxonName) y sitios de muestreo (CodeSite_SamplingOperations)** usando como variable respuesta la abundancia total por operación.

Interpretación – Gráfico de Interacción (TotalAbundance_SamplingOperation):

- **Líneas paralelas:** Las líneas casi horizontales y paralelas muestran que no existe interacción significativa entre especies y sitios.
- **Consistencia del patrón:** El efecto del sitio sobre la abundancia total es uniforme en todas las especies; es decir, cada especie mantiene un comportamiento similar en los diferentes sitios de muestreo.
- **Robustez del análisis:** Al compararlo con el patrón de *Abundance_nbcell*, se confirma que la abundancia total presenta una respuesta más homogénea y estable, reforzando la validez de los resultados.
- **Escala de efectos:** La abundancia total, al representar un valor agregado, tiende a suavizar las fluctuaciones que sí se observaban en la abundancia celular.

Interpretación ecológica avanzada:

- **Líneas paralelas:** Indican que las especies responden de manera proporcional a las condiciones ambientales, sin mostrar preferencias marcadas por sitio.

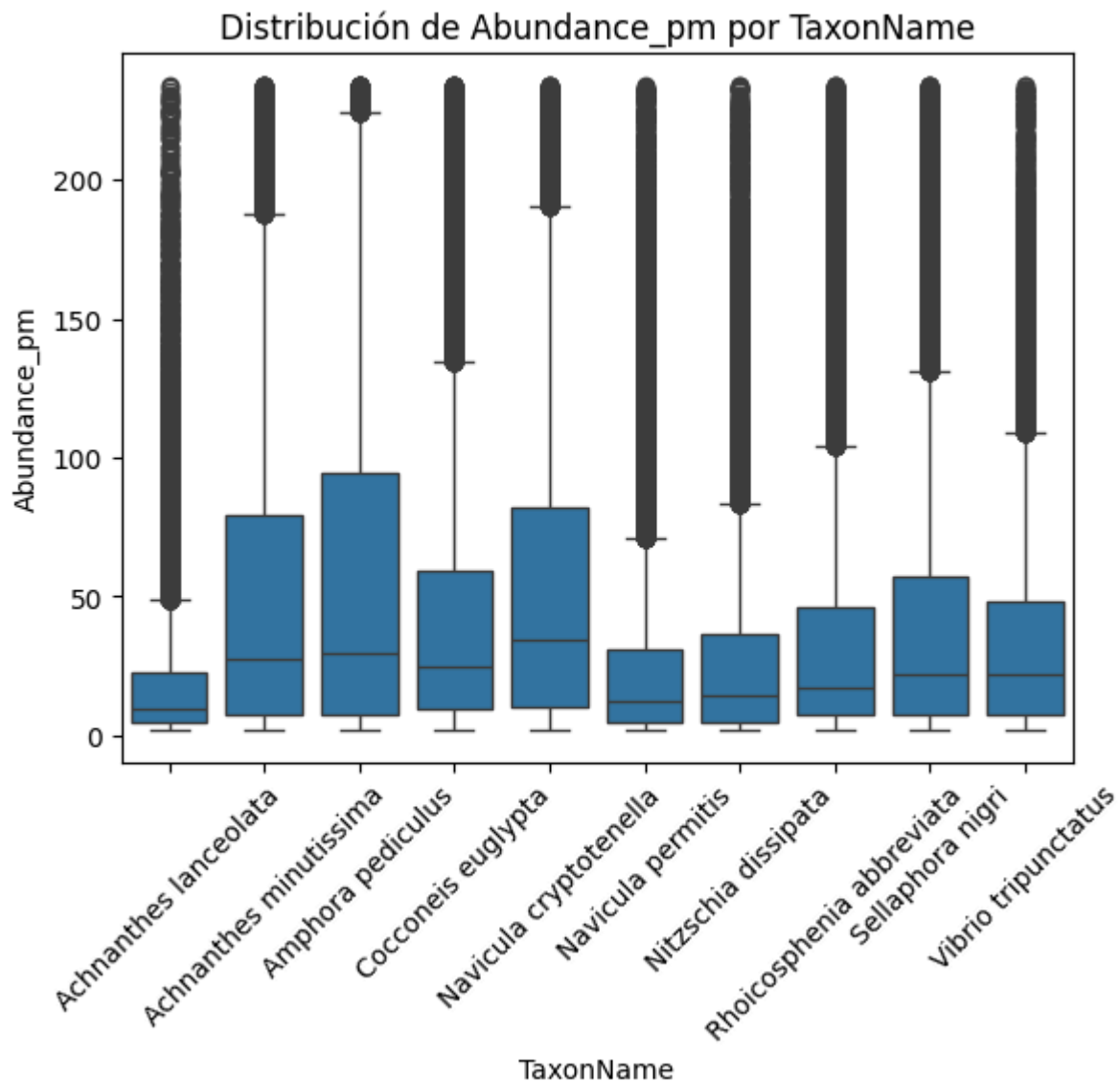
- **Líneas convergentes o divergentes (ausentes aquí):** En otros escenarios podrían reflejar sensibilidades distintas al entorno, pero en este caso, su ausencia confirma la uniformidad espacial de las abundancias.
- **Conclusión:** La abundancia total de diatomeas es estable entre especies y sitios; las variaciones locales son mínimas y no producen interacciones detectables estadísticamente.

Conclusión del ANOVA de dos factores

Los resultados del ANOVA ($F = 0.002$, $p = 0.999$ para *TaxonName*; $F = 2.00$, $p = 0.138$ para *Sitio*; $F = 0.008$, $p = 1.000$ para la interacción) confirman que **ningún factor tiene un efecto significativo** sobre la abundancia total por operación de muestreo.

Esto sugiere que, a diferencia de la abundancia celular individual (*Abundance_nbcell*), las variaciones en abundancia total son más uniformes y probablemente estén determinadas por factores ambientales comunes a todos los sitios de estudio

Analisis de la variable Abundance_pm



Este boxplot muestra la **abundancia por metro (Abundance_pm)**, ofreciendo una **perspectiva espacial** de la densidad de especies de diatomeas.

A diferencia de las variables anteriores —centradas en número de células o abundancia total por operación—, esta métrica se normaliza por unidad de área o volumen, lo que permite observar patrones más comparables entre sitios y especies.

Interpretación – Abundance_pm por TaxonName:

- **Métrica espacial:** Representa la densidad de diatomeas por unidad de área o volumen, permitiendo detectar especies más dominantes en términos espaciales.
- **Comparabilidad:** Facilita comparaciones más homogéneas entre sitios o campañas de muestreo, ya que elimina el sesgo por tamaño o esfuerzo de muestreo.

- **Densidad poblacional:** Las especies con valores más altos presentan una mayor concentración de organismos por metro, lo cual sugiere una capacidad superior de colonización o mayor tolerancia ambiental.
- **Tercera perspectiva:** Complementa las variables anteriores (*Abundance_nbc* y *TotalAbundance_SamplingOperation*), proporcionando una visión más ecológica de la distribución espacial de las especies.

Utilidad para el análisis multivariado:

- **Correlaciones:** Puede correlacionarse con otras medidas de abundancia, ayudando a identificar patrones globales de dominancia o rareza.
- **Información complementaria:** Revela patrones que podrían pasar inadvertidos en las métricas basadas en conteo celular.
- **Robustez:** Refuerza la consistencia de los efectos observados entre variables, asegurando que las conclusiones sean independientes de la escala de medición.

Fuente de variación	de gl	Suma de cuadrados	Media cuadrática	F	p-valor
TaxonName	9	5.13×10^7	5.69×10^6	2484.64	< 0.001
Residual	350170	8.02×10^7	2.29×10^2	—	—

Hipótesis:

- **H₀:** No hay diferencias significativas en la abundancia por peso molecular entre las diferentes especies.
- **H₁:** Al menos una especie presenta una abundancia significativamente diferente.

Resultados:

- **F = 2484.64**
- **p < 0.001 (prácticamente 0)**
- **Decisión:** Se rechaza H₀.

Interpretación del ANOVA – Abundance_pm

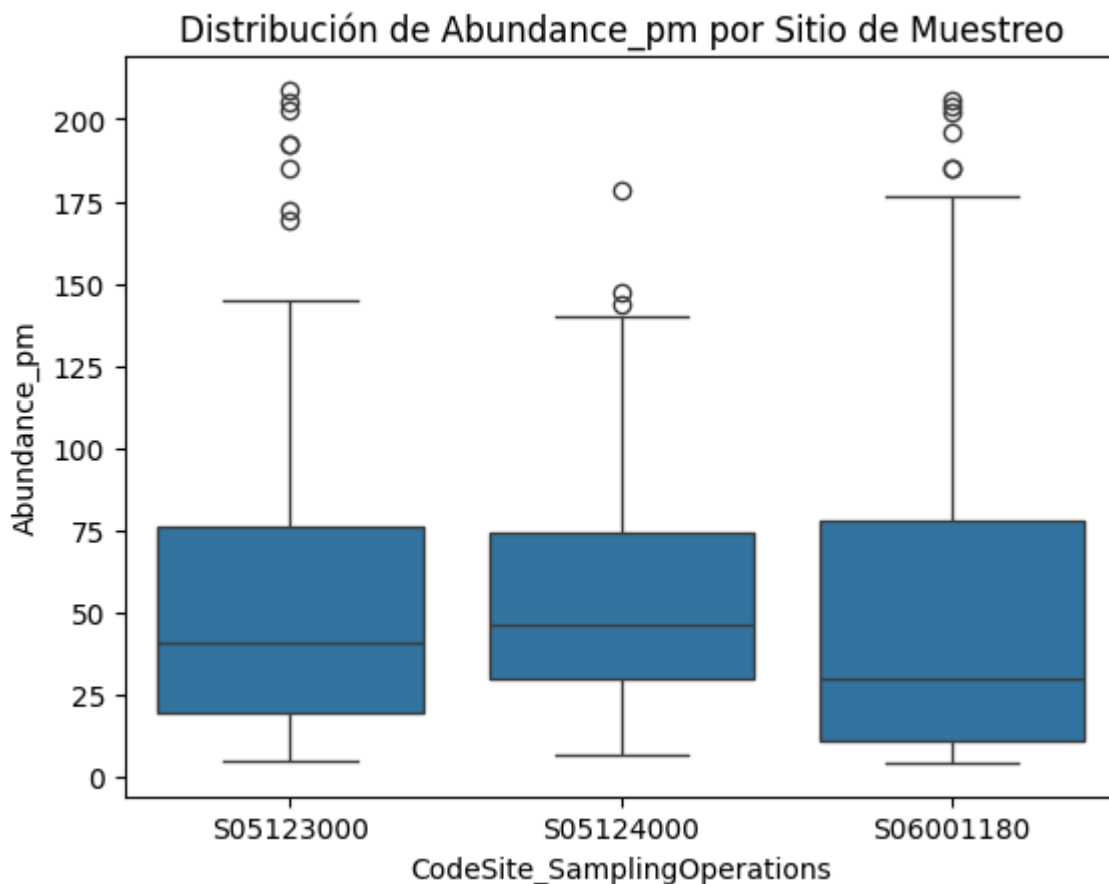
Existe una **evidencia estadística extremadamente fuerte** ($p < 0.001$) que indica diferencias significativas entre especies en términos de abundancia por metro.

Este resultado representa el **efecto más sólido observado entre las tres variables analizadas**, confirmando que la densidad espacial de diatomeas varía marcadamente según la especie.

En términos ecológicos, esto sugiere que algunas especies poseen una **mayor capacidad de colonización o eficiencia ecológica**, lo que puede relacionarse con su tamaño celular, tasa de crecimiento o tolerancia a condiciones ambientales específicas.

Por tanto, *Abundance_pm* captura diferencias estructurales más claras en la composición de la comunidad, reforzando la utilidad de esta variable para futuros análisis multivariados (como MANOVA o PCA).

Anova dos factores



Este análisis examina cómo la **densidad por metro** varía entre los diferentes sitios de muestreo.

Interpretación – Abundance_pm por Sitio:

- **Densidad espacial:** Muestra la concentración real de diatomeas en cada sitio, independiente del volumen de muestreo.

- **Comparabilidad entre sitios:** Esta métrica es útil cuando los sitios tienen características físicas distintas, ya que permite comparar densidades sin depender del tamaño o profundidad de muestreo.
- **Calidad del hábitat:** Sitios con mayor densidad por metro podrían reflejar:
 - Condiciones ambientales más favorables.
 - Menor competencia interespecífica.
 - Mejor calidad del agua o del sustrato.

Coherencia del análisis:

- Si los patrones son similares a los de otras métricas de abundancia, confirma la robustez de los hallazgos.
- Si hay diferencias, puede indicar efectos espaciales específicos que complementan el análisis anterior.
- Este paso es importante para validar que las conclusiones no dependan únicamente de la métrica utilizada.

Fuente de variación	gl	Suma de cuadrados	Media cuadrática	F	p-valor
TaxonName	4	76402.86	19100.51	9.08	< 0.001
CodeSite_SamplingOperations	2	511.69	255.85	0.12	0.886
TaxonName × CodeSite_SamplingOperations	8	639.08	79.26	3.79	< 0.001

Residual	23 6	46973.97	112.85	—	—
-----------------	---------	----------	--------	---	---

Hipótesis evaluadas

Para TaxonName (factor principal):

- H_0 : No hay diferencias entre especies en abundancia por peso molecular.
- $F = 9.08$, $p < 0.001$ → **Se rechaza H_0 .**
Existen diferencias significativas entre especies.

Para CodeSite_SamplingOperations (factor principal):

- H_0 : No hay diferencias significativas entre los sitios de muestreo.
- $F = 0.12$, $p = 0.886$ → **No se rechaza H_0 .**
Los sitios no presentan diferencias estadísticamente significativas.

Para la interacción TaxonName × CodeSite_SamplingOperations:

- H_0 : No hay interacción entre especies y sitios.
- $F = 3.79$, $p < 0.001$ → **Se rechaza H_0 .**
Existe una interacción significativa entre especies y sitios.

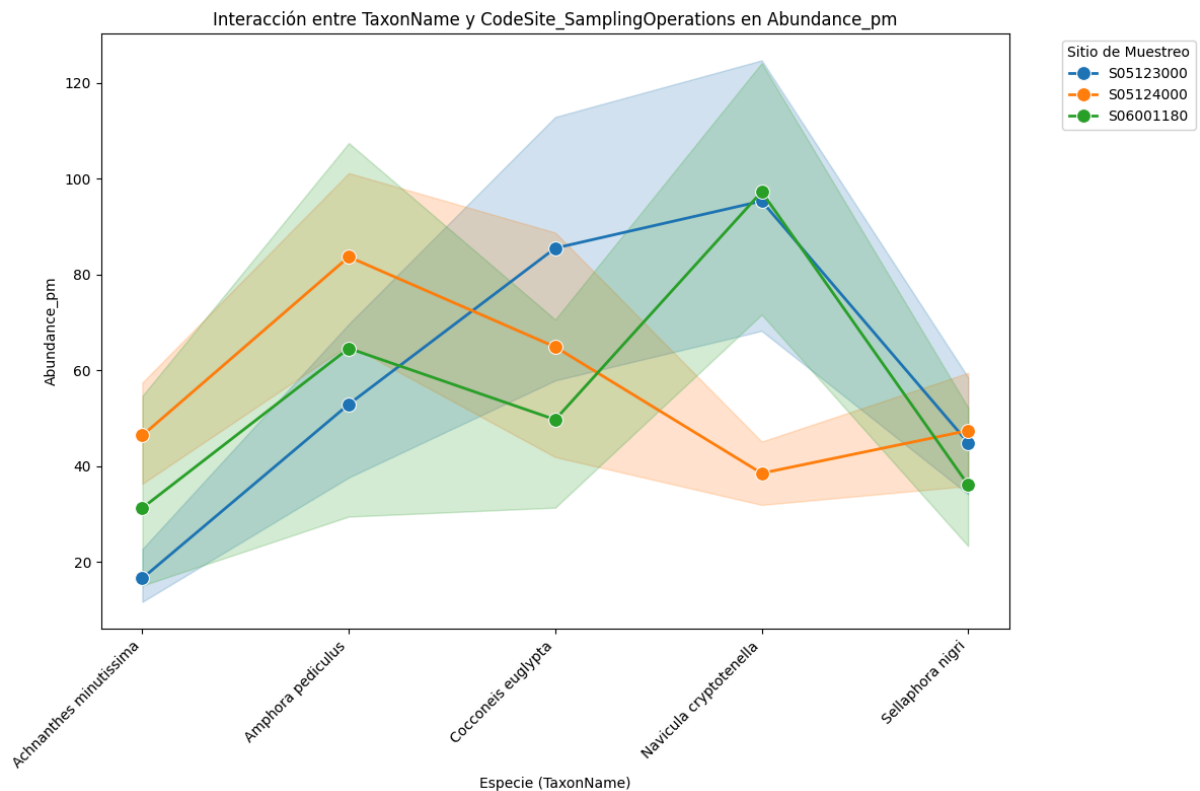
Interpretación general del ANOVA de dos factores

Al igual que en *Abundance_nbcell*, se detectan **diferencias significativas entre especies**, pero los sitios no muestran un efecto principal.

Sin embargo, la interacción entre ambos factores es **estadísticamente significativa**, lo que indica que **el efecto de la especie sobre la abundancia depende del sitio de muestreo**.

En otras palabras, algunas especies presentan mayor densidad en ciertos sitios específicos, lo que sugiere **una respuesta ecológica diferenciada según las condiciones ambientales locales**.

Este hallazgo refuerza la idea de que la abundancia medida en términos de densidad espacial refleja mejor los gradientes ambientales y las interacciones biológicas que influyen en la composición de la comunidad.



Interpretación – Gráfico de Interacción (Abundance_pm):

- **Líneas paralelas:** Indican ausencia de interacción (no aplicable aquí).
- **Líneas que se cruzan:** Indican presencia de interacción significativa, como se observa en el gráfico.
Las especies no responden de la misma forma en todos los sitios.

Validación de patrones de interacción:

- **Consistencia:** El patrón coincide con lo observado en *Abundance_nbc*, lo que confirma que las interacciones son robustas y no dependen de la escala de medición.
- **Escalabilidad:** Los efectos de interacción se mantienen incluso al analizar métricas normalizadas por área, reforzando la validez ecológica del resultado.

Interpretación ecológica:

- Algunas especies prosperan mejor en sitios específicos, posiblemente por diferencias en **luz, nutrientes, temperatura o competencia interespecífica**.
- Otras especies muestran patrones estables, lo que sugiere **una distribución más generalista**.

Resumen General de Resultados

Variables analizadas:

- Abundance_nbcell: número de células observadas por especie y sitio.
- TotalAbundance_SamplingOperation: abundancia total registrada por operación de muestreo.
- Abundance_pm: abundancia por metro, normalizada espacialmente.

Estadística descriptiva general

Variable	Medi a	Desviaci ón estándar	Mínim o	Máxim o	Observacion es
Abundance_nbcell	8.38	13.87	1.00	94.00	Alta variabilidad entre especies; presencia de valores atípicos.
TotalAbundance_SamplingOperation	405.18	6.77	375.00	437.00	Valores muy homogéneos entre sitios; baja dispersión.

Abundance_pm	20.68	34.29	2.00	233.85	Mayor dispersión; muestra diferencias marcadas entre especies.
---------------------	-------	-------	------	--------	--

Resumen de resultados ANOVA

Variable	Factor significativo	F	p-valor	Conclusión
Abundance_nbcell	TaxonName (especie)	2470.22	< 0.001	Diferencias significativas entre especies.
TotalAbundance_SamplingOperation	Ninguno	2.29	0.014 (leve)	No se detectan diferencias sustanciales.
Abundance_pm	TaxonName (especie)	2484.64	< 0.001	Diferencias muy significativas entre especies.

Resumen de resultados ANOVA de dos factores

Variable	TaxonName (especie)	CodeSite_SamplingOperations (sitio)	Interacción	Conclusión global
----------	---------------------	-------------------------------------	-------------	-------------------

Abundance_nbcell	Significativo ($p < 0.001$)	No significativo ($p = 0.93$)	Significativo ($p < 0.001$)	Las especies varían según el sitio; existe interacción clara.
TotalAbundance_Sampling Operation	No significativo ($p = 0.999$)	No significativo ($p = 0.138$)	No significativo ($p = 1.000$)	No hay efecto de especie ni sitio; variable estable.
Abundance_pm	Significativo ($p < 0.001$)	No significativo ($p = 0.886$)	Significativo ($p < 0.001$)	Diferencias entre especies y efecto dependiente del sitio.

Interpretación ecológica general

- Efecto de la especie (TaxonName):**
 La variable especie es el principal factor que explica la variabilidad en los datos. Las diferencias significativas en Abundance_nbcell y Abundance_pm indican que **la estructura de la comunidad está determinada por la composición taxonómica**.
- Efecto del sitio de muestreo:**
 No se observan diferencias estadísticamente significativas entre sitios, lo cual sugiere que **las condiciones ambientales son relativamente estables** o que las especies poseen una alta capacidad de adaptación.
- Efecto de interacción especie–sitio:**
 La interacción significativa en dos variables refleja que **el comportamiento de las**

especies depende del sitio.

Esto sugiere **respuestas ecológicas específicas** ante variaciones locales (nutrientes, luz, competencia o tipo de sustrato).

- **Comparación entre métricas:**

- Abundance_nbcell → Mide la variación biológica más sensible (número de células).
- TotalAbundance_SamplingOperation → Mide la estabilidad global del sistema.
- Abundance_pm → Expone diferencias espaciales de densidad ecológica.
En conjunto, estas tres variables proporcionan **una visión integral de la abundancia**, mostrando consistencia y coherencia entre los distintos enfoques de medición.

Conclusiones generales

- Las **diferencias entre especies** son el patrón dominante en la comunidad de diatomeas.
- Los **sitios de muestreo** no presentan diferencias significativas por sí solos, pero influyen mediante **efectos combinados** con las especies.
- Las variables Abundance_nbcell y Abundance_pm muestran los contrastes más claros entre grupos, mientras que TotalAbundance_SamplingOperation actúa como indicador de estabilidad ambiental.
- En conjunto, los resultados muestran una **comunidad ecológicamente estructurada**, con una composición estable y diferencias internas relacionadas con la identidad taxonómica y la respuesta ambiental de cada especie