

ANOVA Y ANOVA DE DOS FACTORES

TEAM MEMBERS (DATAFORGE)



**JESÚS EDUARDO VALLE
VILLEGAS**

Finanzas
A01770616



**DIEGO ANTONIO OROPEZA
LINARTE**

BGB
A01733018



**MANUEL EDUARDO
COVARRUBIAS RODRÍGUEZ**

ITC
A01737781



**ITHANDEHUI JOSELYN
ESPINOZA**

ITC
A01734547



**MAURICIO GRAU GUTIERREZ
RUBIO**

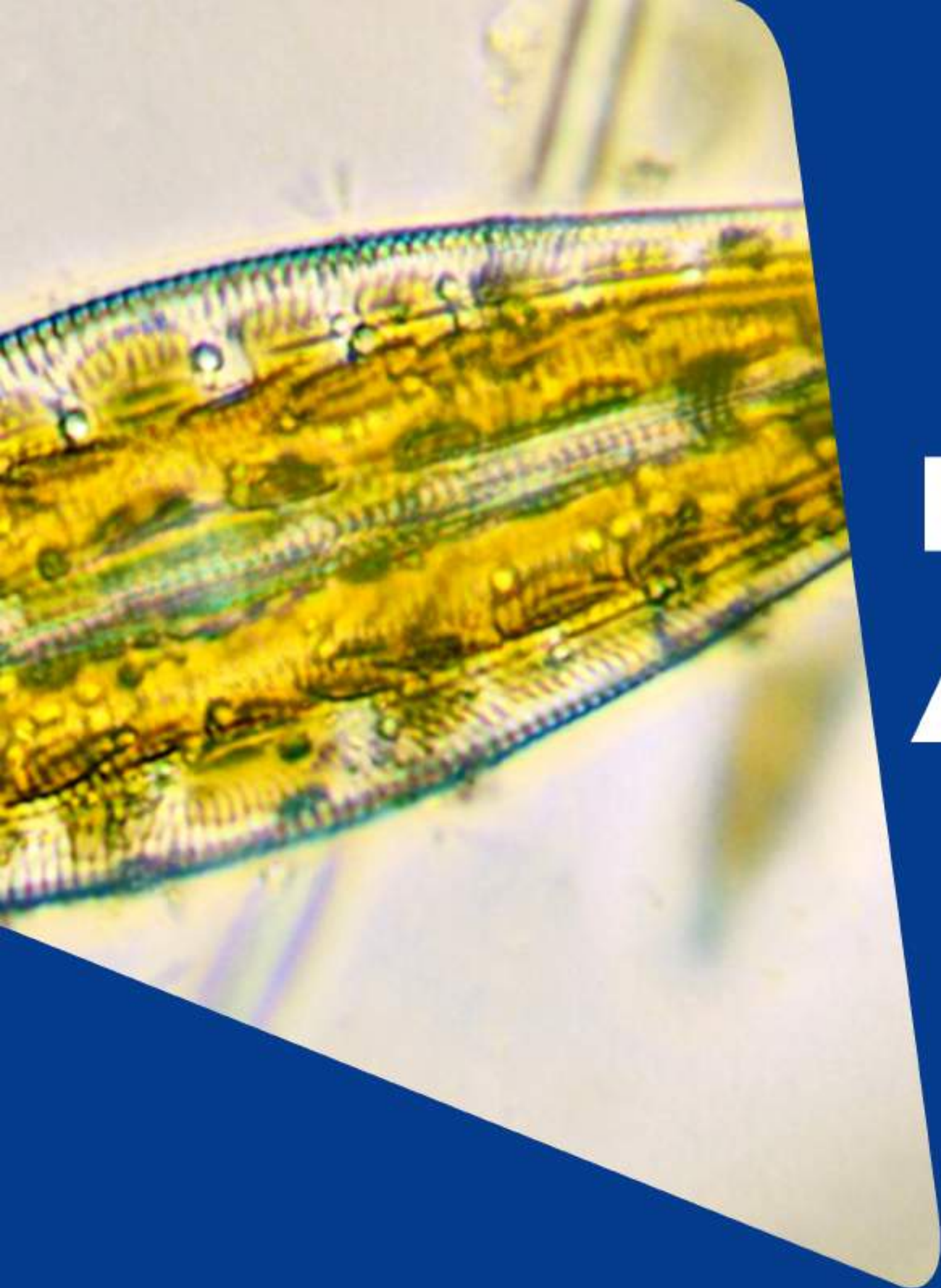
LEM
A01734914

Objetivo general:

Aplicar métodos estadísticos, como ANOVA de uno y dos factores, para descubrir si existen diferencias importantes entre grupos mediante el p-value, y analizar qué tan bien funcionan distintos modelos usando el F1-score.

Metodología que usó el team DataForge



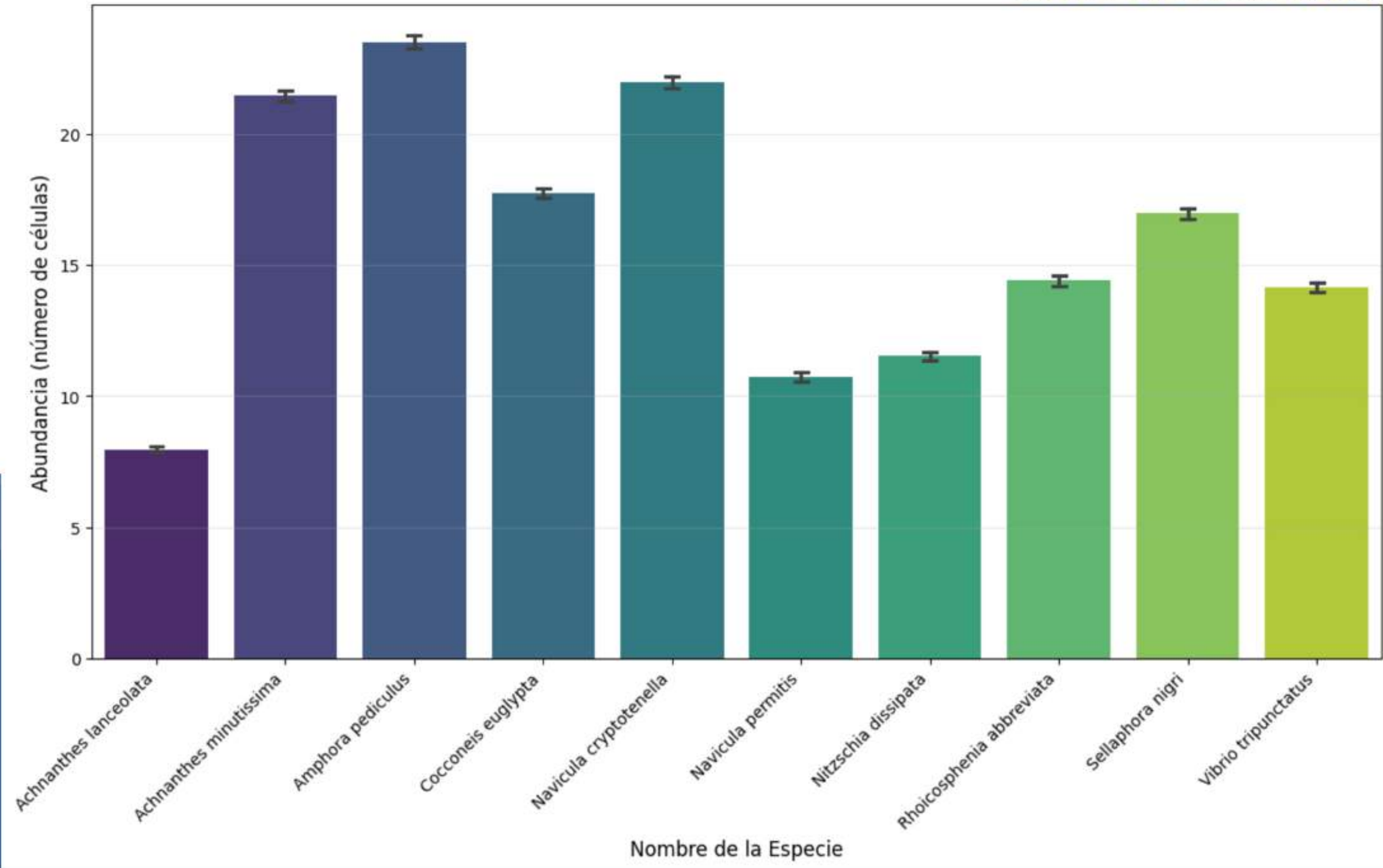


Datathón

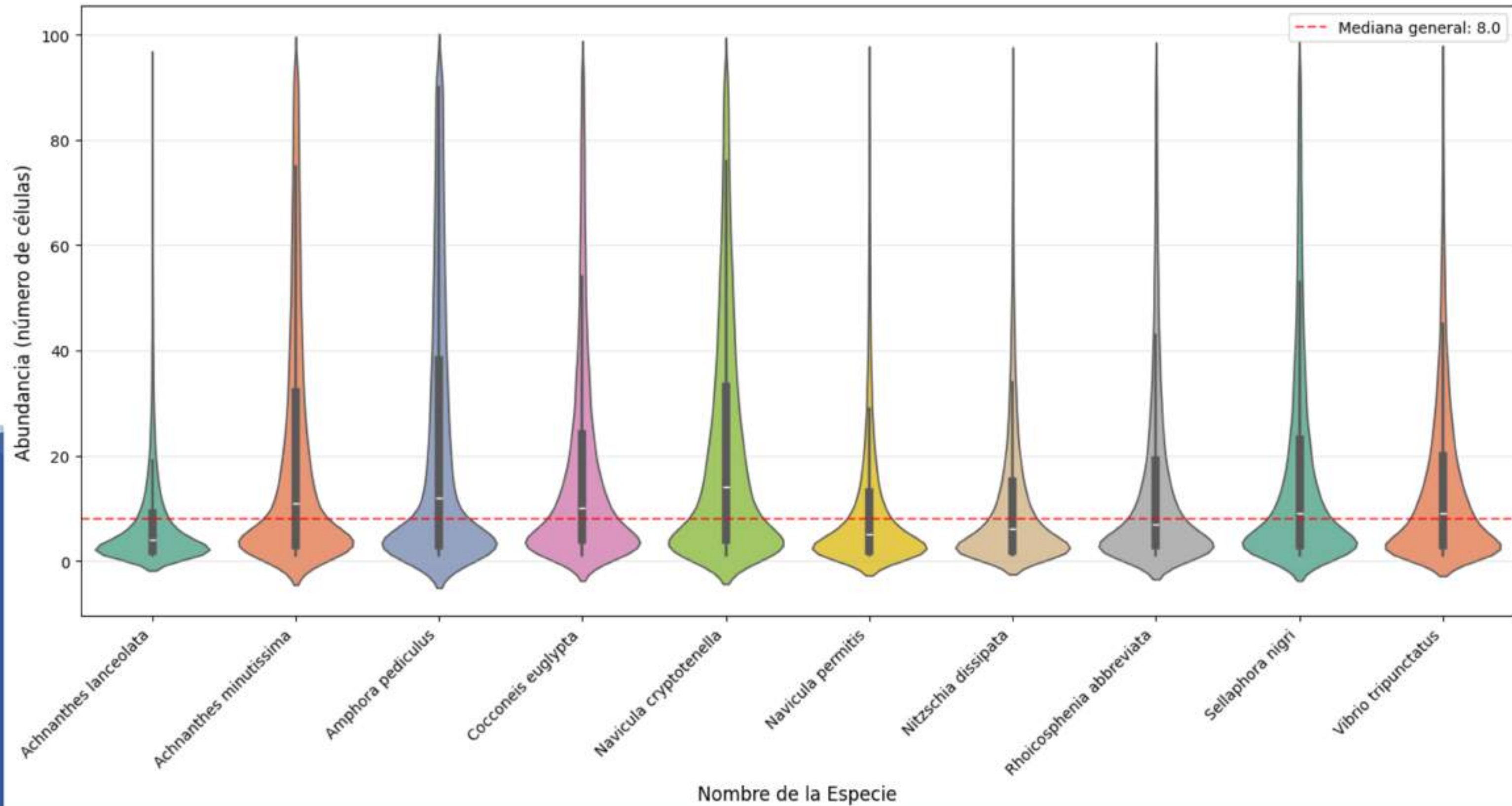
Actividad 5.1

Abundancia Promedio por Especie (Top 10)

Intervalos de confianza (95%)



Variabilidad en la abundancia celular



NATURALEZA DE LAS VARIABLES INDEPENDIENTES

TaxonName y CodeSite_SamplingOperations

Elemento	Descripción
Variables Explicativas	Son variables categóricas que representan grupos o factores
SIGNIFICADO BIOLÓGICO TaxonName (Especies)	Pregunta científica: ¿Las diferentes especies tienen abundancias diferentes?
CodeSite_SamplingOperations (Sitios de Muestreo)	<ul style="list-style-type: none">• Pregunta científica: ¿El lugar de muestreo afecta la abundancia? - condiciones ambientales distintas (temperatura, pH, nutrientes, luz)
OTRAS VARIABLES NO UTILIZADAS COMO FACTORES	<ul style="list-style-type: none">• TaxonCode: Es redundante con TaxonName (misma información)• SamplingOperations_code: Es más un identificador técnico que un factor biológico• Date_SamplingOperation: Aunque se podría usar, el análisis se centró en especies y sitios como factores principales



Resultados del ANOVA de un solo factor (Abundance_nbcell)

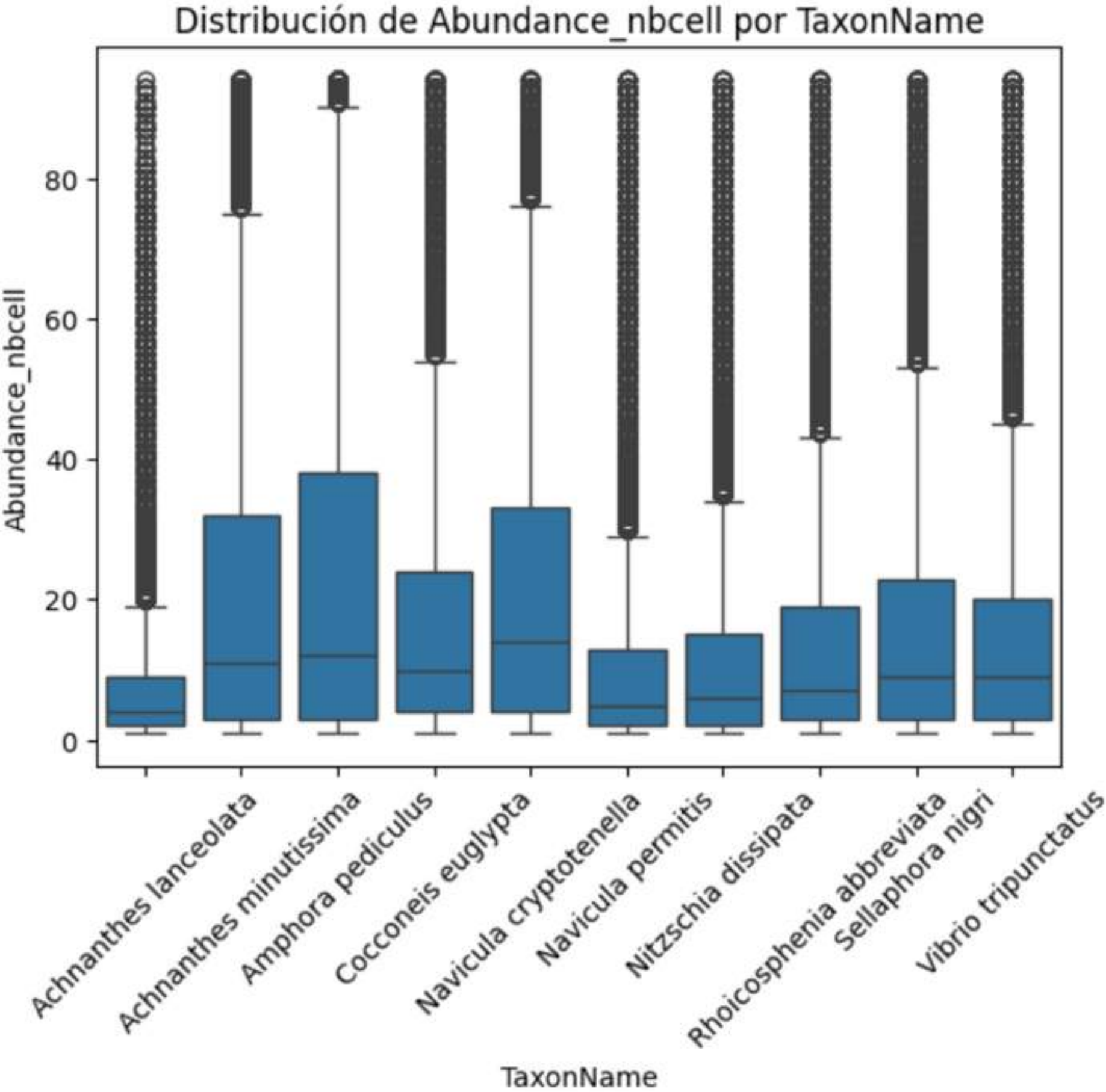
vs TaxonName

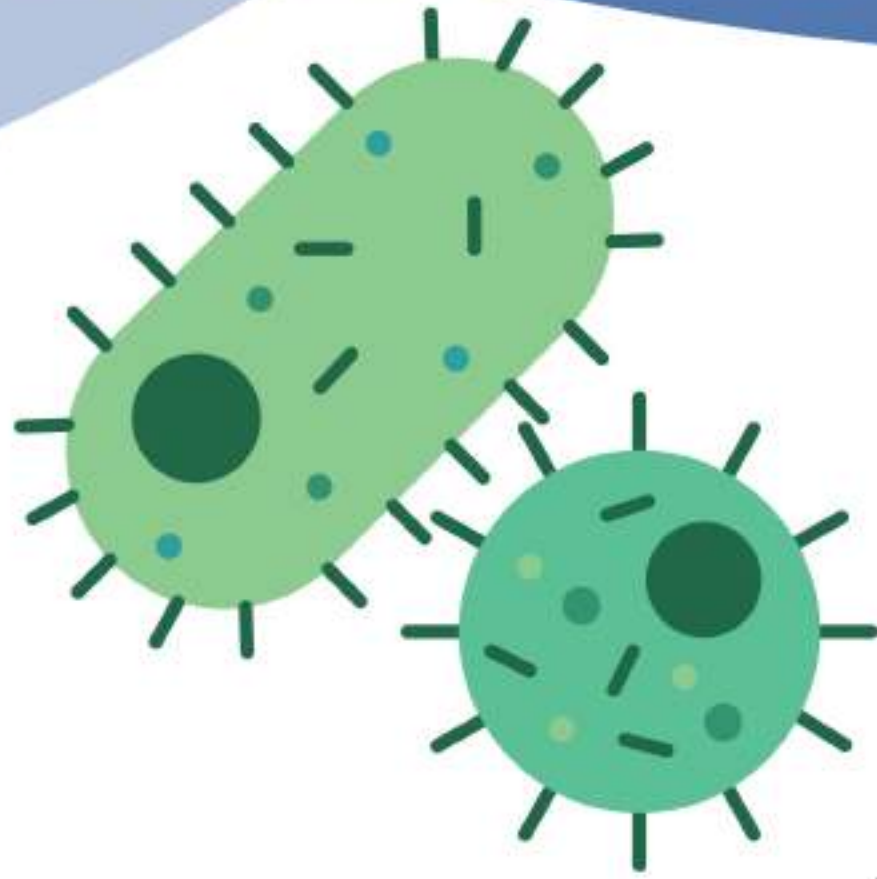
- H_0 : No hay diferencias significativas en la abundancia de células entre las diferentes especies.
- H_1 : Al menos una especie tiene una abundancia de células significativamente diferente.

Fuente de variación	F	p-value (PR(>F))
TaxonName	2,470.22	< 0.001

Decisión: RECHAZAMOS H_0

Las diferentes especies de diatomeas presentan abundancias de células significativamente diferentes. Las especies no tienen la misma abundancia promedio de células.



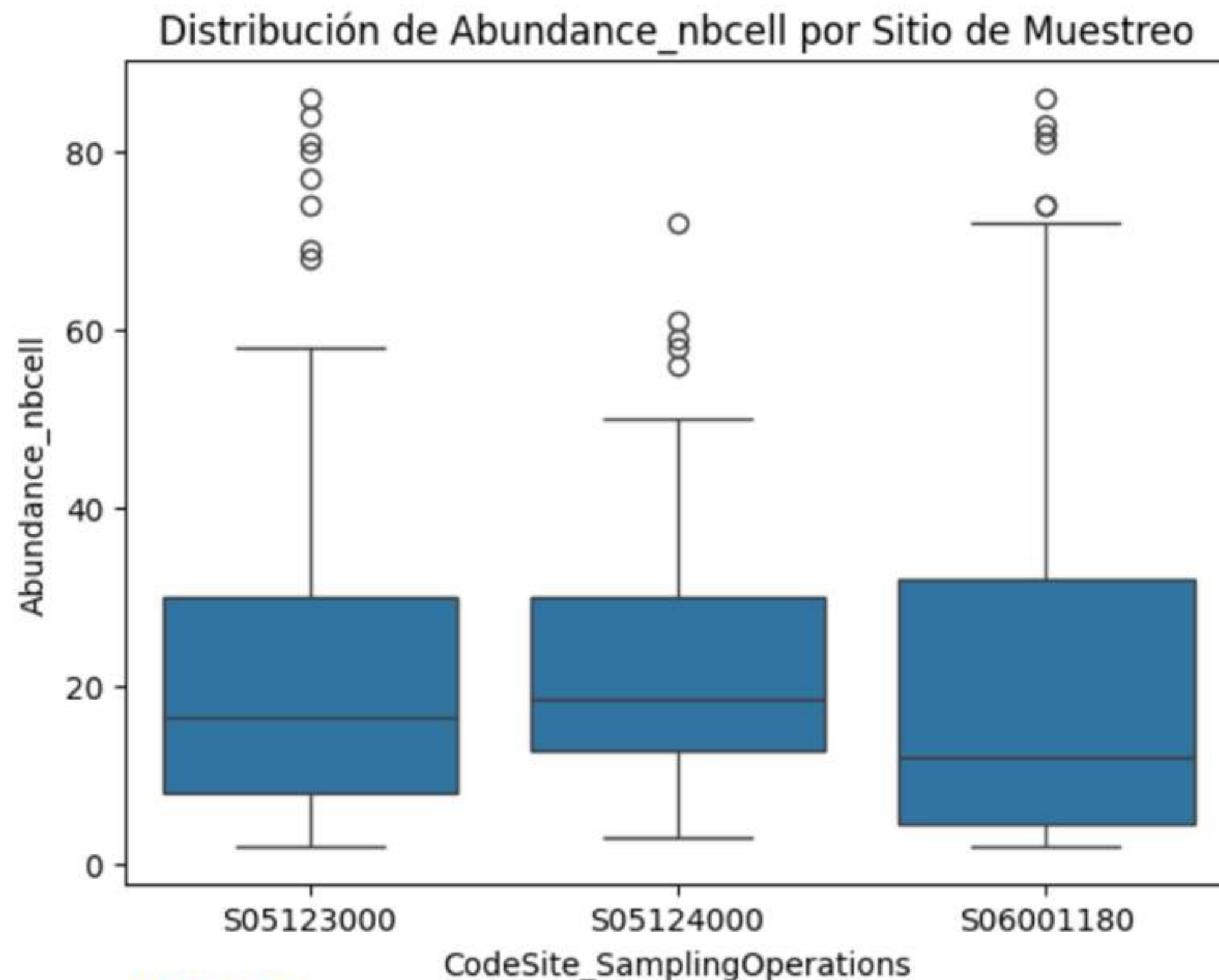


Resultados del ANOVA de dos factores (Abundance_nbcell)

vs TaxonName + CodeSite_SamplingOperations +
TaxonName:CodeSite_SamplingOperations

Fuente de variación	F	p-value (PR(>F))
TaxonName	9.14	< 0.001
CodeSite_SamplingOperations	0.07	0.932
TaxonName × CodeSite_SamplingOperations	3.79	< 0.001

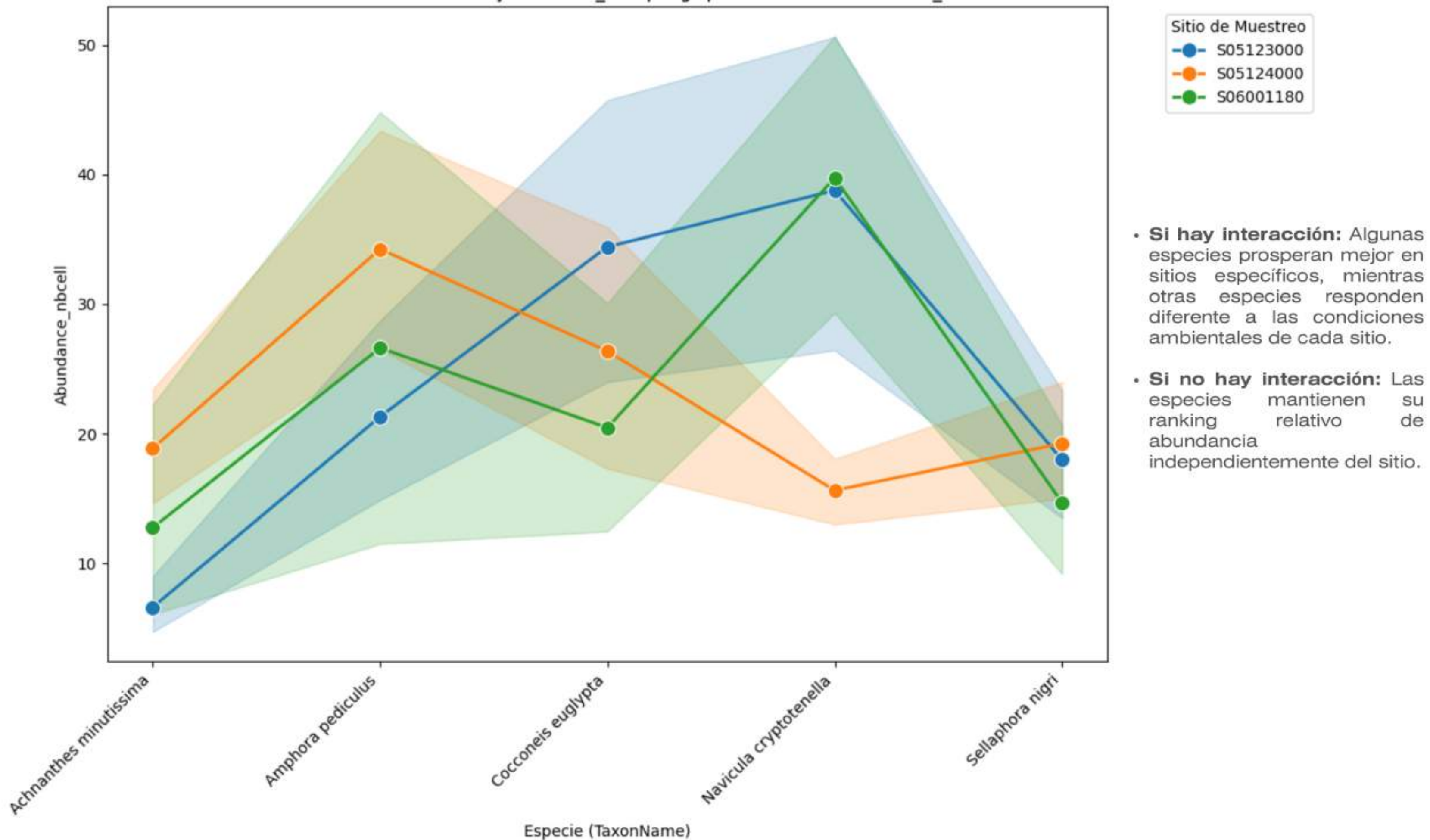
- Las especies difieren significativamente en abundancia.
- Los sitios por sí solos NO tienen efecto significativo.
- Sí existe interacción significativa entre especies y sitios (las especies responden diferente según el sitio).

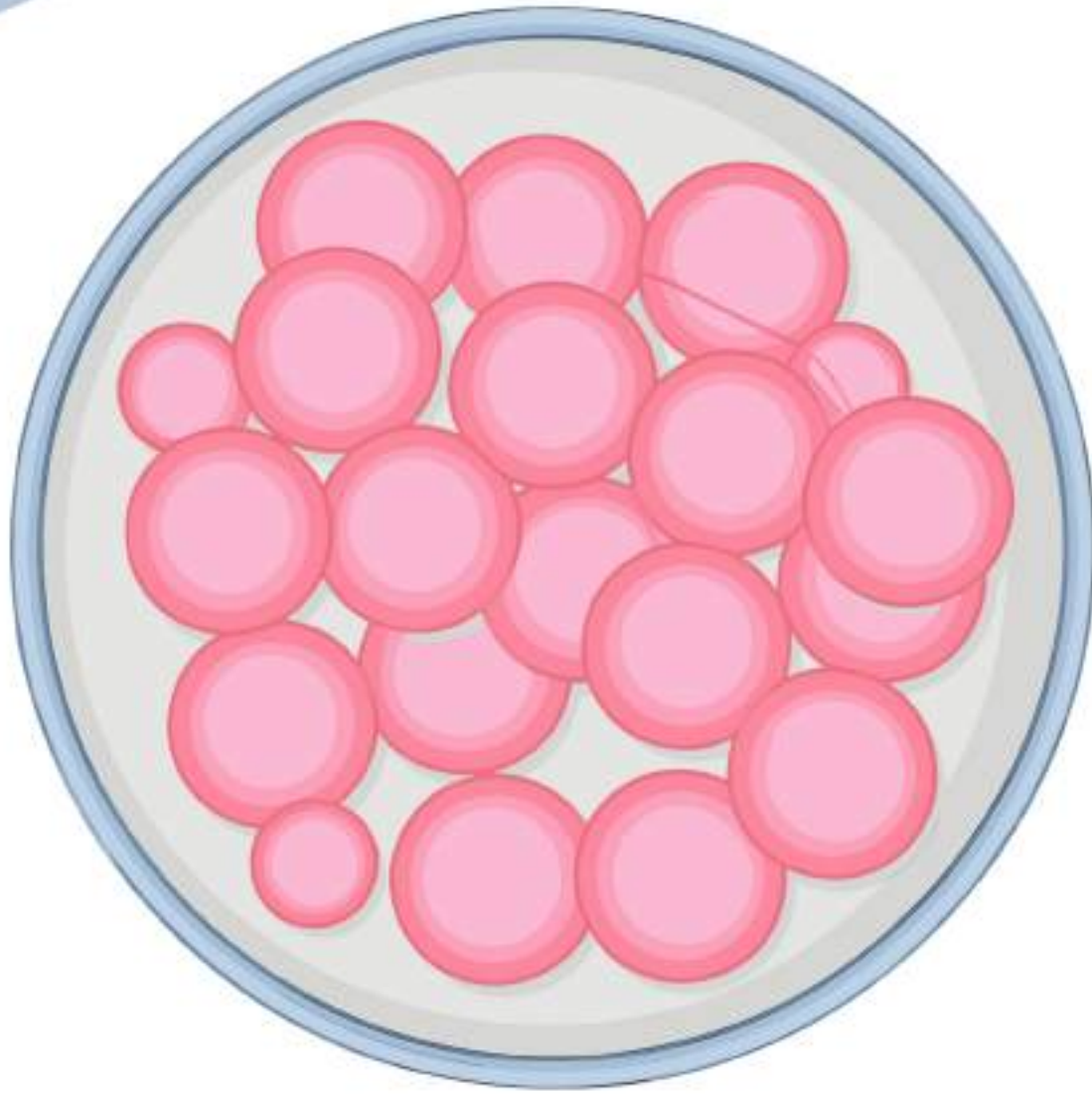


Decisión:

- Para TaxonName: **RECHAZAMOS H_0**
- Para CodeSite_SamplingOperations: **NO RECHAZAMOS H_0**
- Para Interacción TaxonName:CodeSite_SamplingOperations: **RECHAZAMOS H_0**

Interacción entre TaxonName y CodeSite_SamplingOperations en Abundance_nbccl





Resultados del ANOVA de un solo factor (TotalAbundance_SamplingOperation)

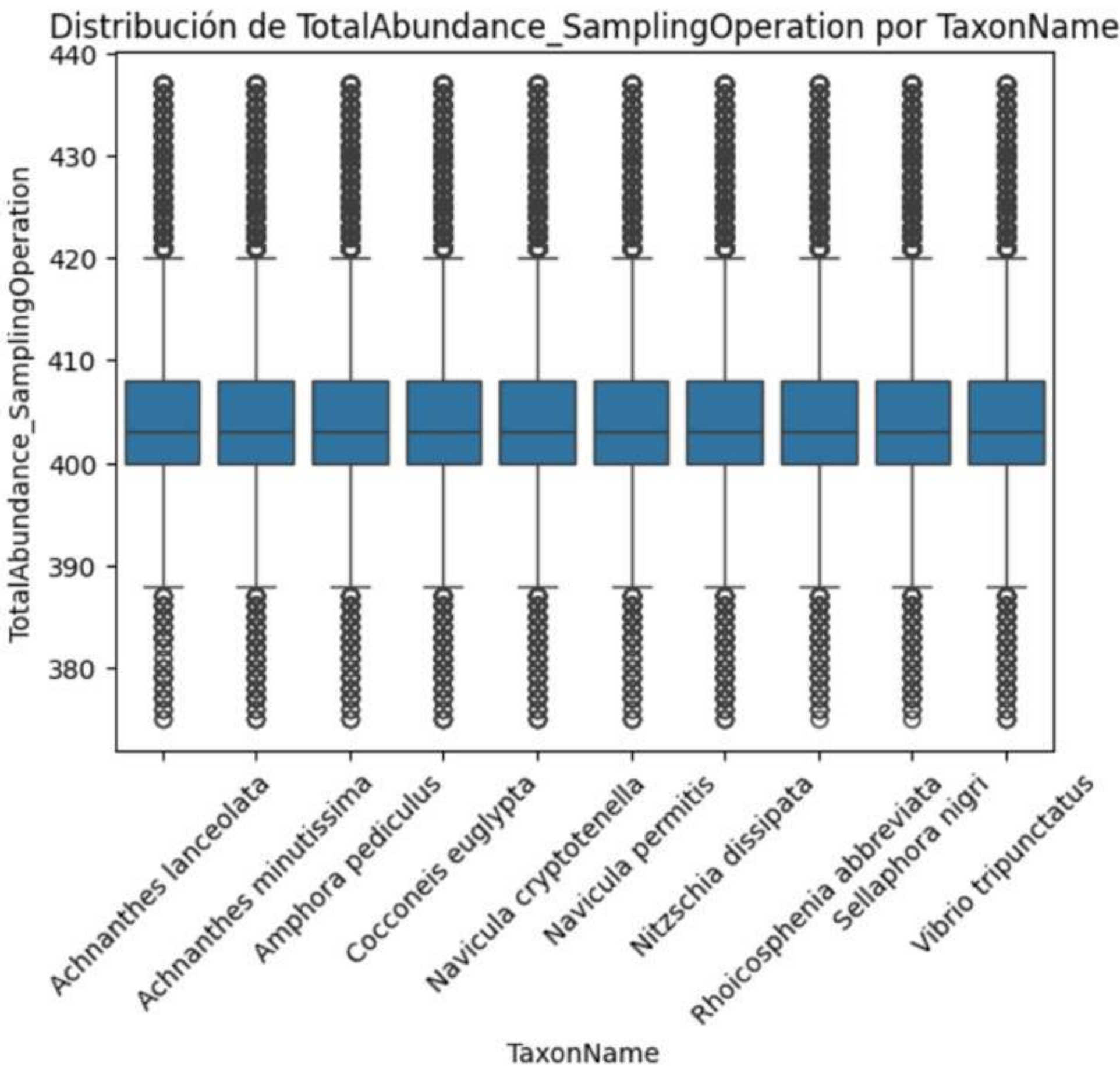
vs TaxonName

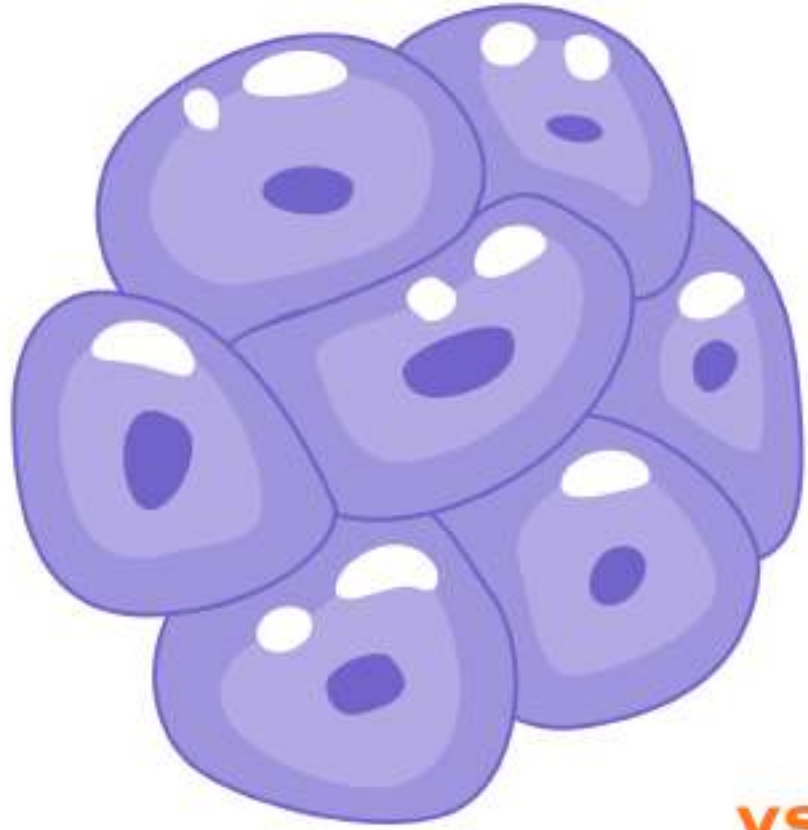
- H_0 : No hay diferencias significativas en la abundancia total por operación de muestreo entre las diferentes especies.
- H_1 : Al menos una especie tiene una abundancia total por operación de muestreo significativamente diferente.

Fuente de variación	F	p-value (PR(>F))
TaxonName	2.29	0.0143

Decisión: RECHAZAMOS H_0

Las diferentes especies de diatomeas presentan abundancias totales por operación de muestreo significativamente diferentes.





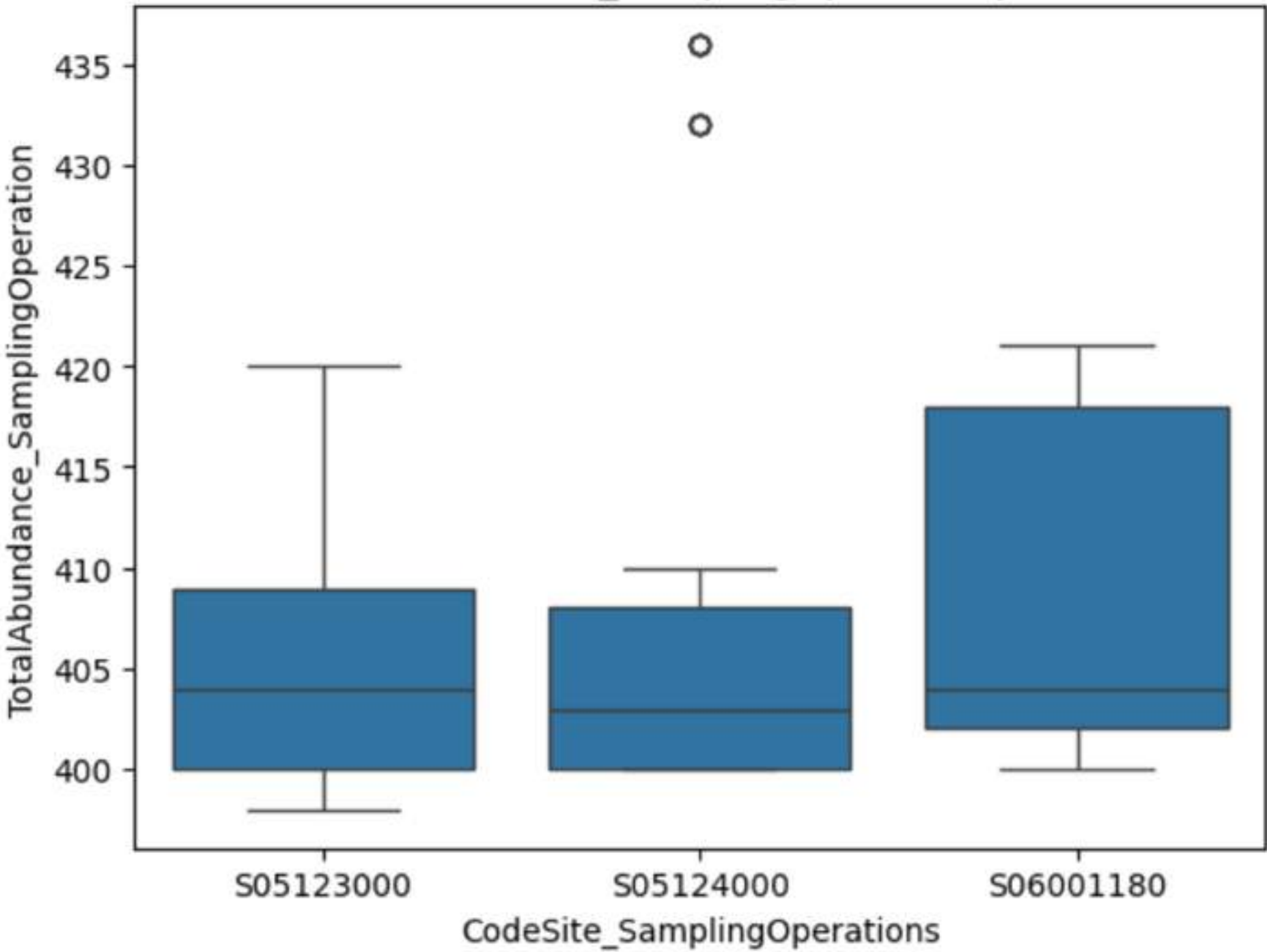
Resultados del ANOVA de dos factores (TotalAbundance_ SamplingOperation)

vs TaxonName + CodeSite_SamplingOperations +
TaxonName:CodeSite_SamplingOperations

Fuente de variación	F	p-value (PR(>F))
TaxonName	0.002	0.999
CodeSite_SamplingOperations	2.00	0.138
TaxonName × CodeSite_SamplingOperations	0.008	1.000

- Ni las especies, ni los sitios, ni su interacción muestran diferencias significativas en la abundancia total por operación de muestreo cuando se controlan ambos factores simultáneamente.

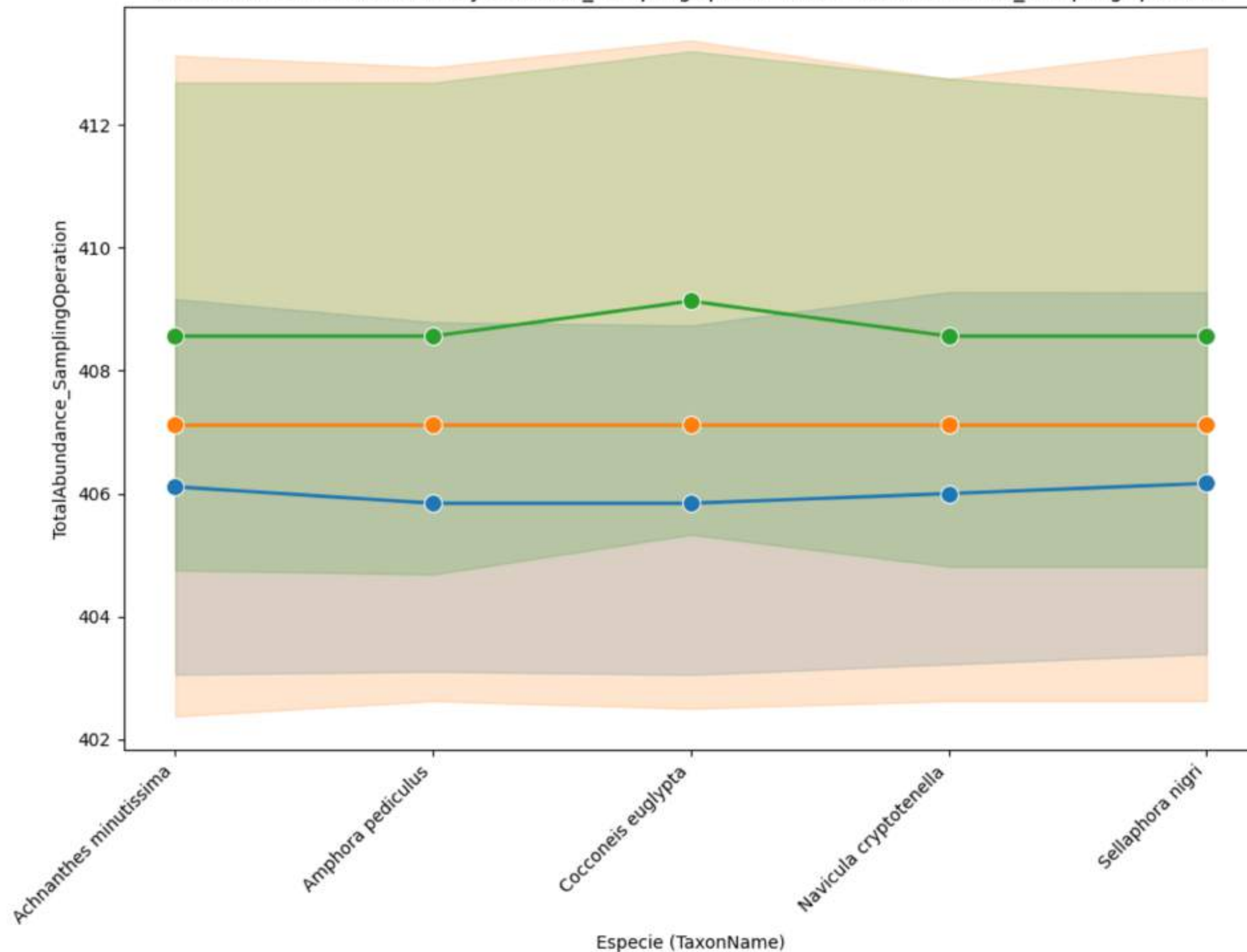
Distribución de TotalAbundance_SamplingOperation por Sitio de Muestreo



Decisión:

- Para TaxonName: **NO RECHAZAMOS H_0**
- Para CodeSite_SamplingOperations: **NO RECHAZAMOS H_0**
- Para Interacción TaxonName:CodeSite_SamplingOperations: **NO RECHAZAMOS H_0**

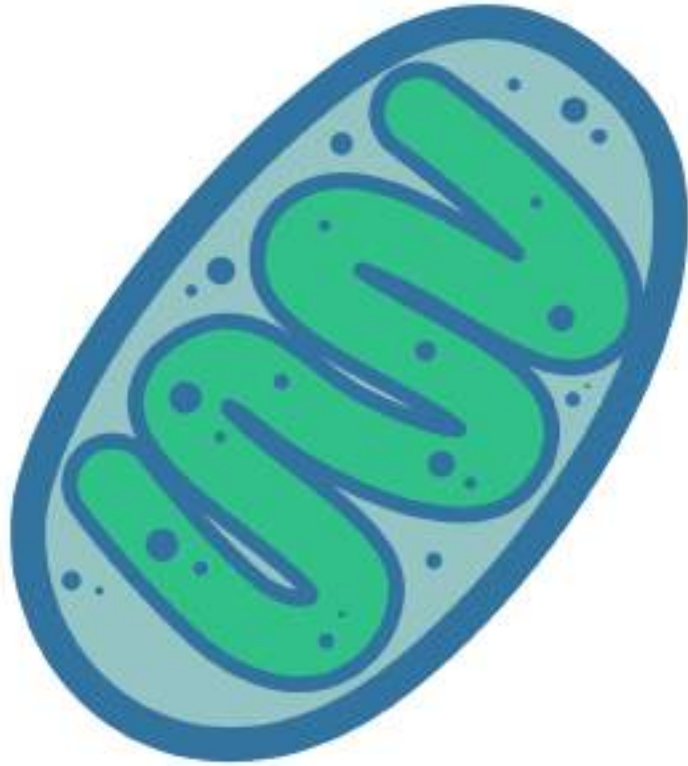
Interacción entre TaxonName y CodeSite_SamplingOperations en TotalAbundance_SamplingOperation



Sitio de Muestreo

- S05123000
- S05124000
- S06001180

- **Líneas paralelas:** Las especies responden de manera proporcional a las condiciones de cada sitio.
- **Líneas convergentes/divergentes:** Algunas especies son más sensibles a las diferencias ambientales entre sitios.
- **Líneas que se cruzan:** Existe una interacción fuerte donde especies diferentes dominan en sitios diferentes.



Resultados del ANOVA de un solo factor (Abundance_pm)

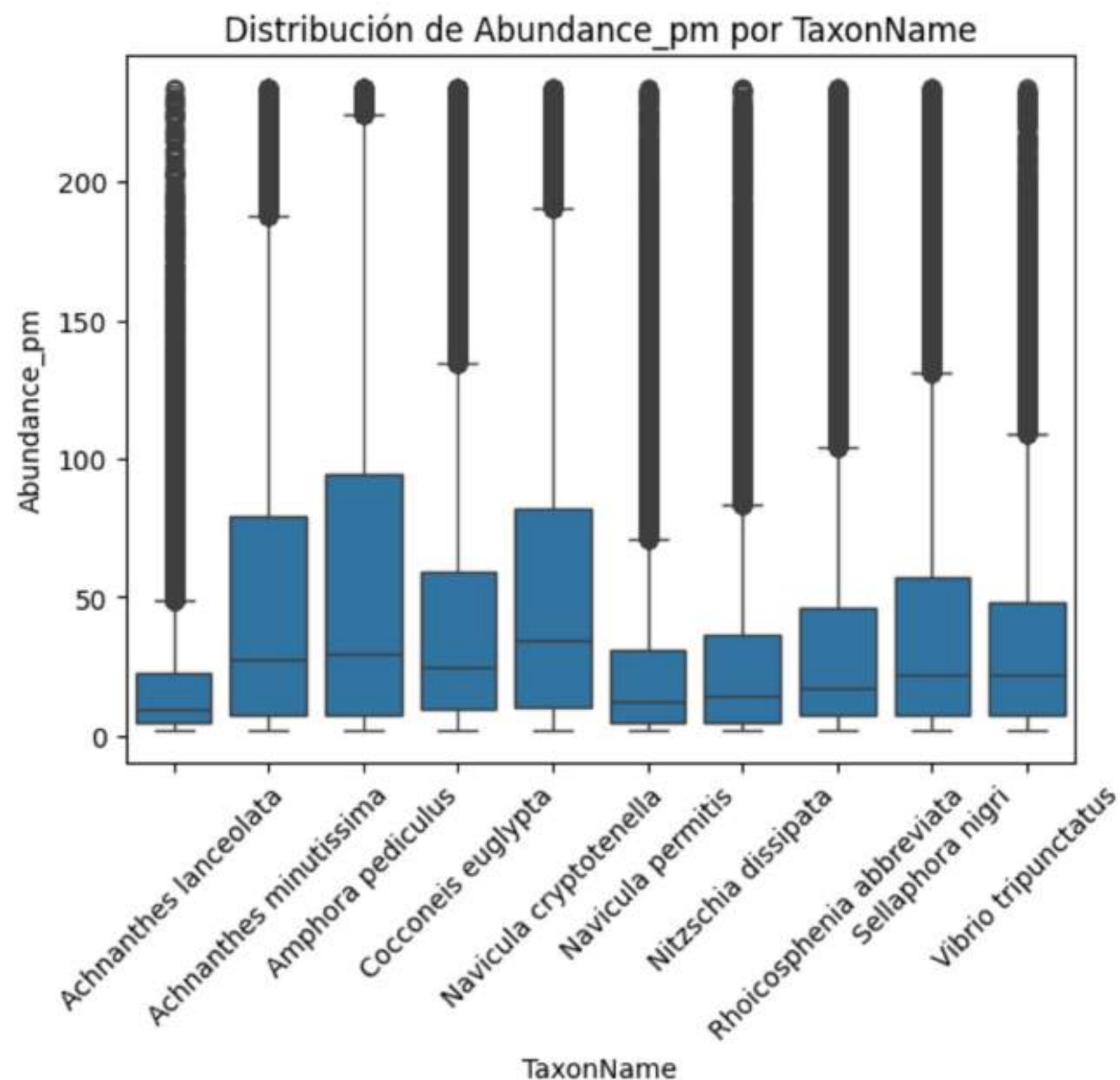
vs TaxonName

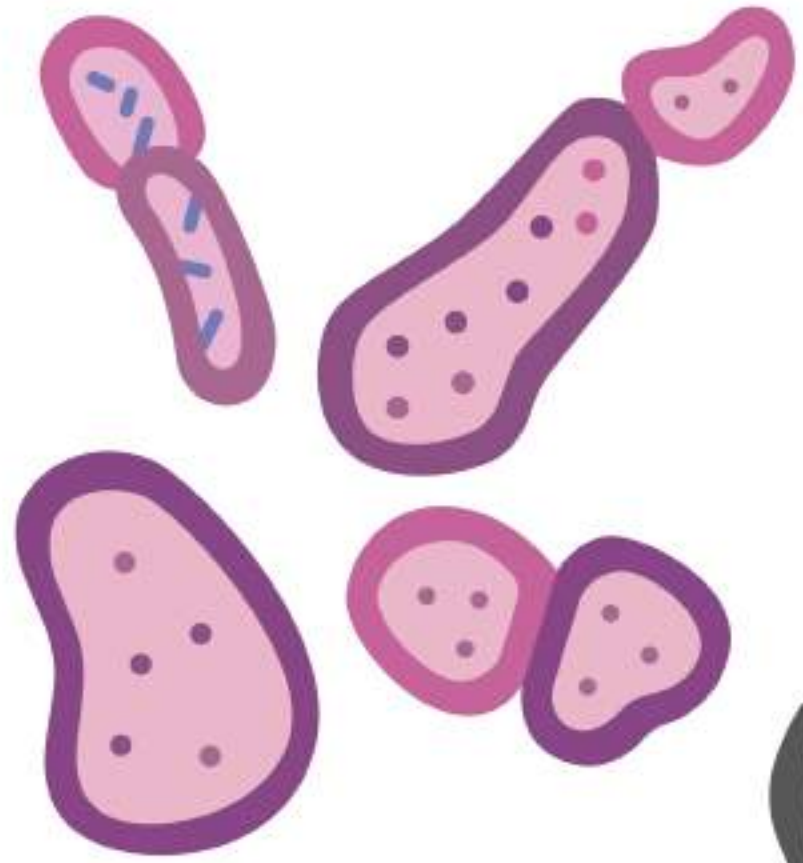
- H_0 : No hay diferencias significativas en la abundancia por peso molecular entre las diferentes especies.
- H_1 : Al menos una especie tiene una abundancia por peso molecular significativamente diferente.

Fuente de variación	F	p-value (PR(>F))
TaxonName	2,484.64	< 0.001

Decisión: RECHAZAMOS H_0

Las diferentes especies de diatomeas presentan abundancias por peso molecular significativamente diferentes.



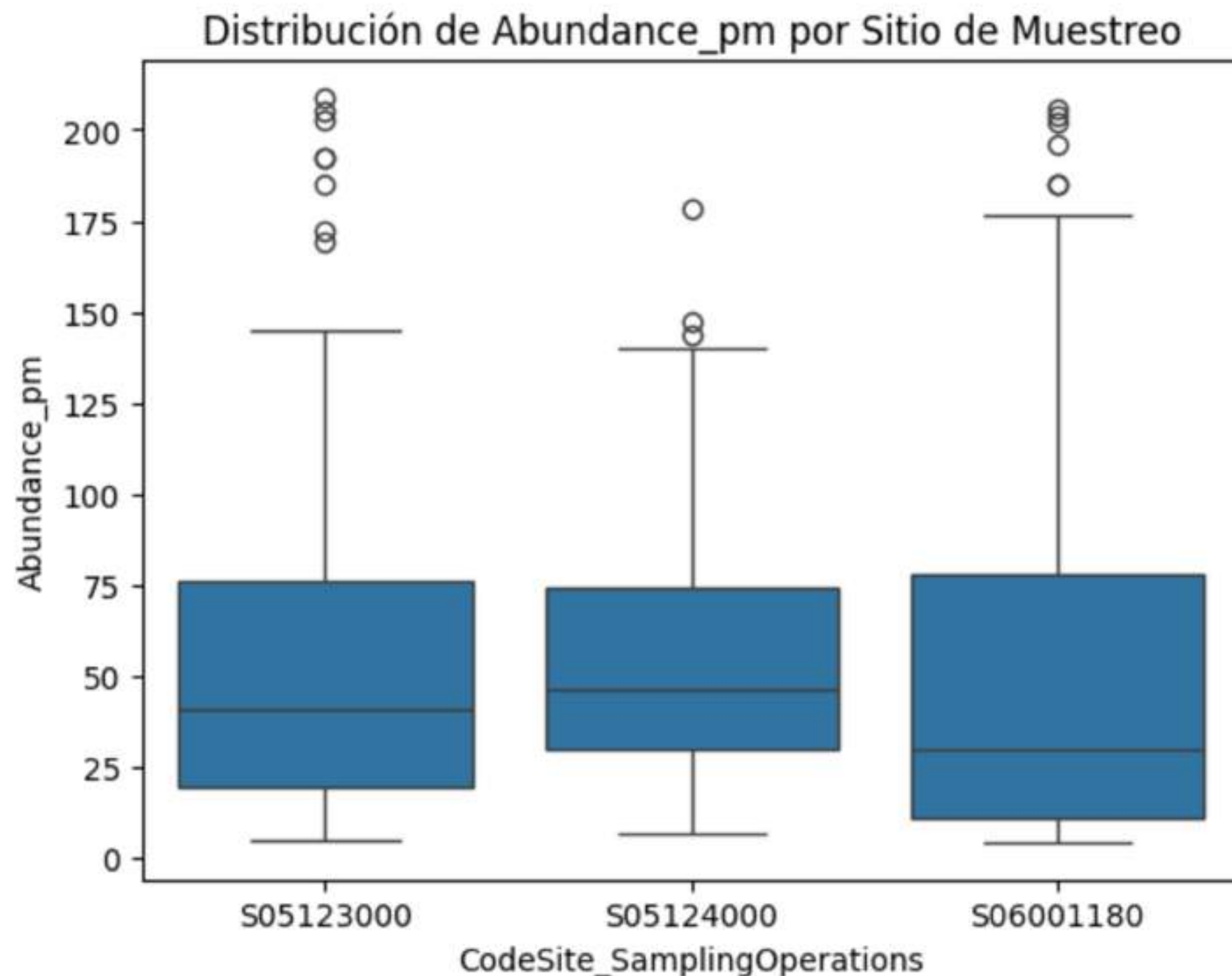


Resultados del ANOVA de dos factores (Abundance_pm)

vs TaxonName + CodeSite_SamplingOperations +
TaxonName:CodeSite_SamplingOperations

Fuente de variación	F	p-value (PR(>F))
TaxonName	9.08	< 0.001
CodeSite_SamplingOperations	0.12	0.886
TaxonName × CodeSite_SamplingOperations	3.79	< 0.001

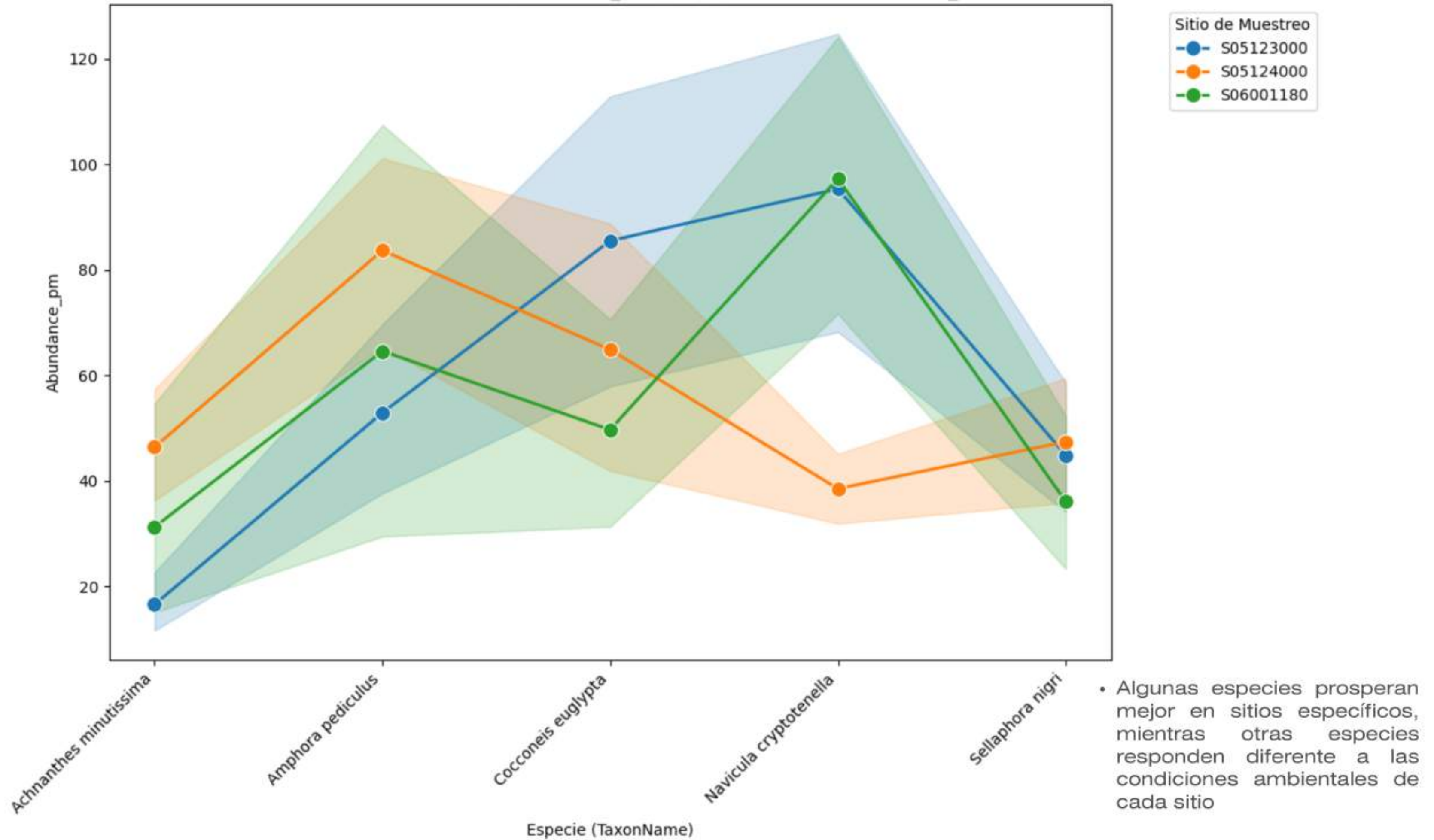
- Similar al caso de Abundance_nbcell.
- Las especies difieren significativamente.
- los sitios NO tienen efecto principal.
- Sí hay interacción significativa entre especies y sitios.
- El efecto de las especies depende del sitio de muestreo.



Decisión:

- Para TaxonName: **RECHAZAMOS H_0**
- Para CodeSite_SamplingOperations: **NO RECHAZAMOS H_0**
- Para Interacción TaxonName:CodeSite_SamplingOperations: **RECHAZAMOS H_0**

Interacción entre TaxonName y CodeSite_SamplingOperations en Abundance_pm



ANÁLISIS DE INSIGHTS

ANOVA (Un Factor)
– Evaluar
diferencias entre
especies

**ANOVA (Dos
Factores)** – Evaluar
efectos principales
e interacciones

**Sí existe
interacción
sitio×especie**

**Las especies tienen
patrones de
abundancia
distintivos**

**El sitio por sí solo
NO es
determinante, pero
Sí interactúa con
las especies**



FORVIA

Actividad 5.2



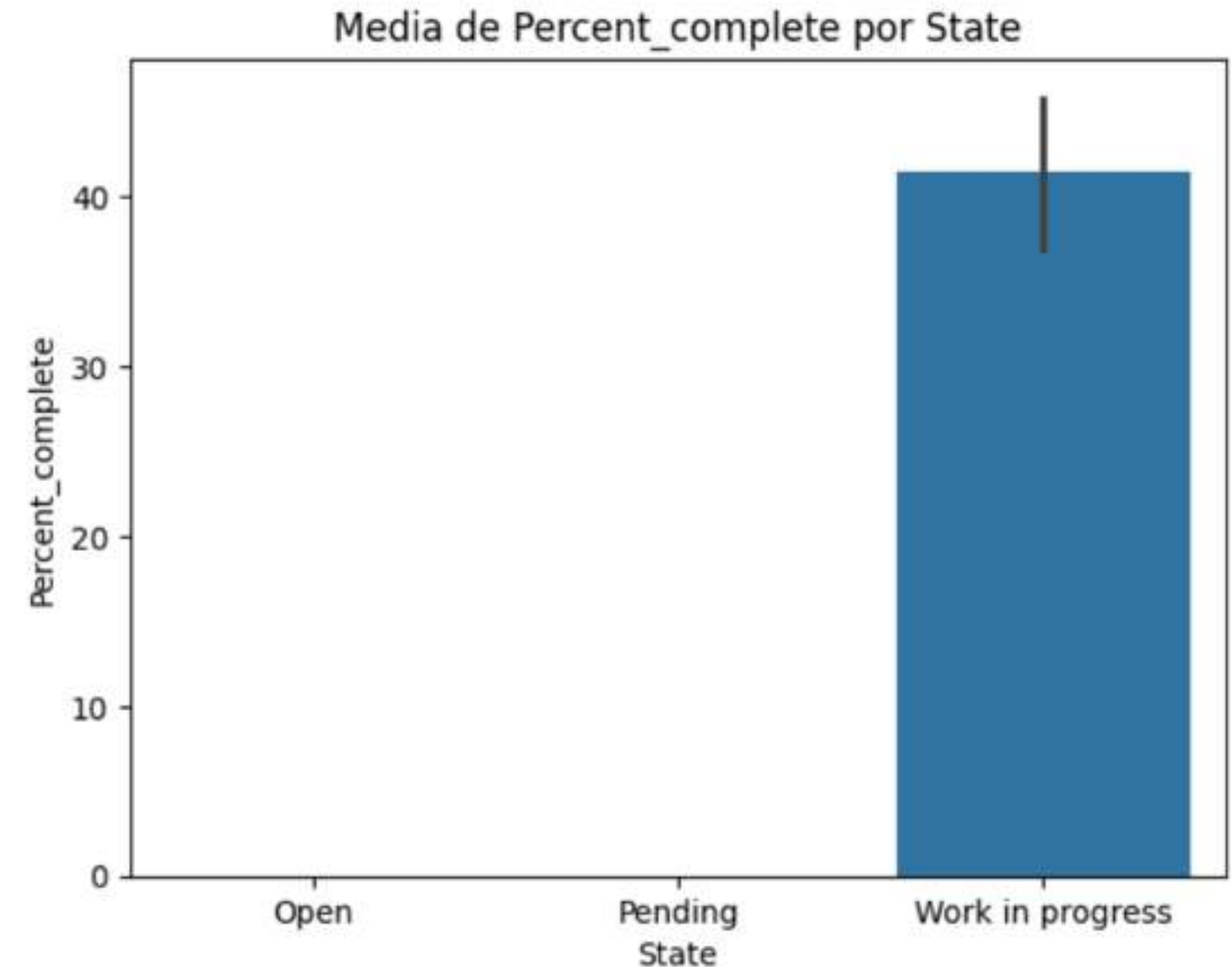
Percent_complete por State

Decisión:

Se rechaza H_0 , ya que $p < 0.05$.

Existen diferencias significativas en el porcentaje de avance entre los estados del proyecto.

Fuente de variación	F	p-value (PR(>F))
State	9.04	0.00016



Los proyectos en progreso ("Work in progress") presentan un porcentaje de avance significativamente mayor que los proyectos abiertos ("Open") o pendientes ("Pending"), lo que confirma que el estado operativo influye directamente en el avance del proyecto.

Percent_complete por State

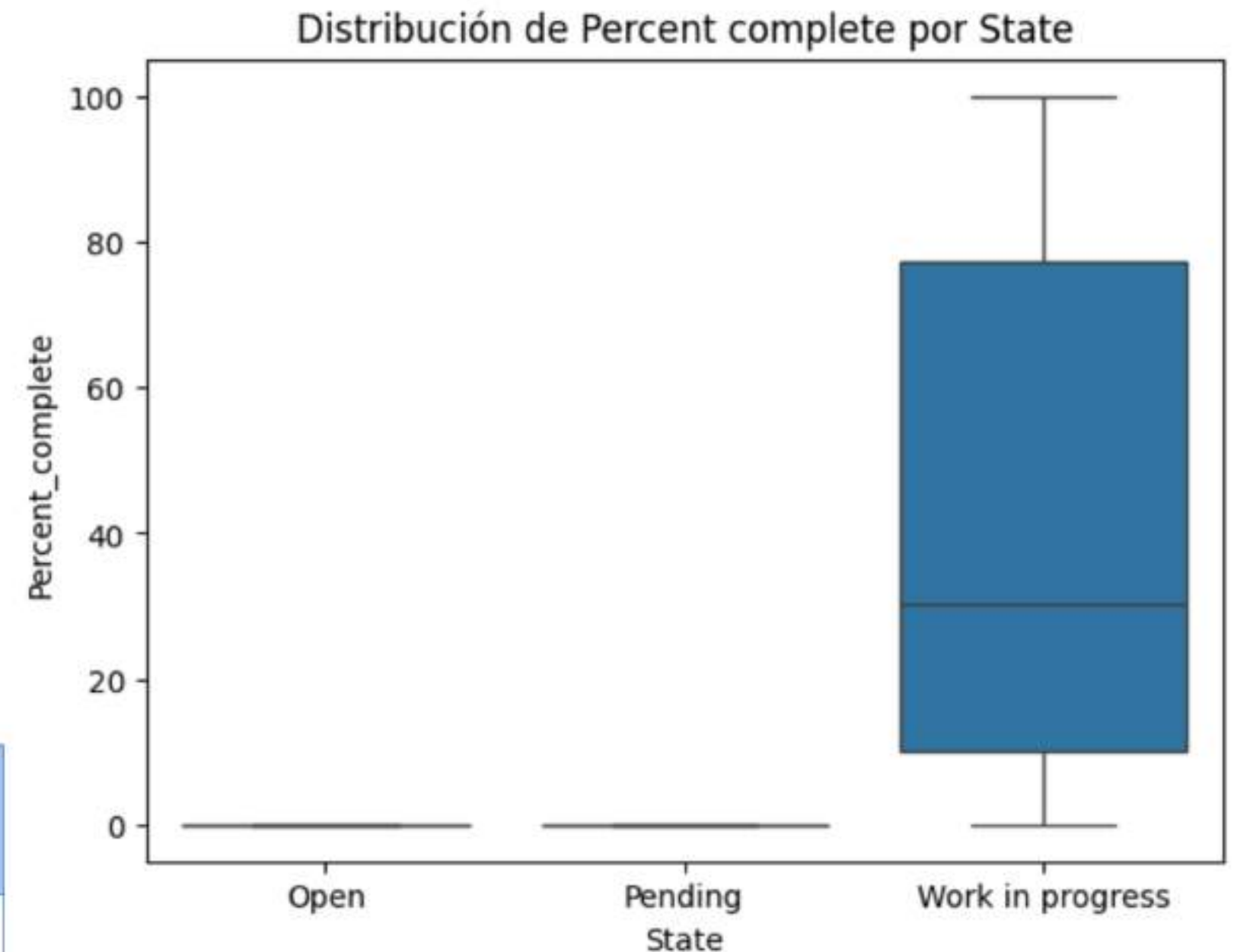
Se rechaza H_0 solo para la variable State ($p < 0.05$).

Los factores Project_Type y la interacción State \times Project_Type no presentan diferencias significativas.

Interpretación:

El porcentaje de avance (Percent_complete) depende del estado del proyecto, pero no del tipo de proyecto, ni de la combinación entre ambos factores.

Fuente de variación	F	p-value (PR(>F))
State	9.04	0.00016

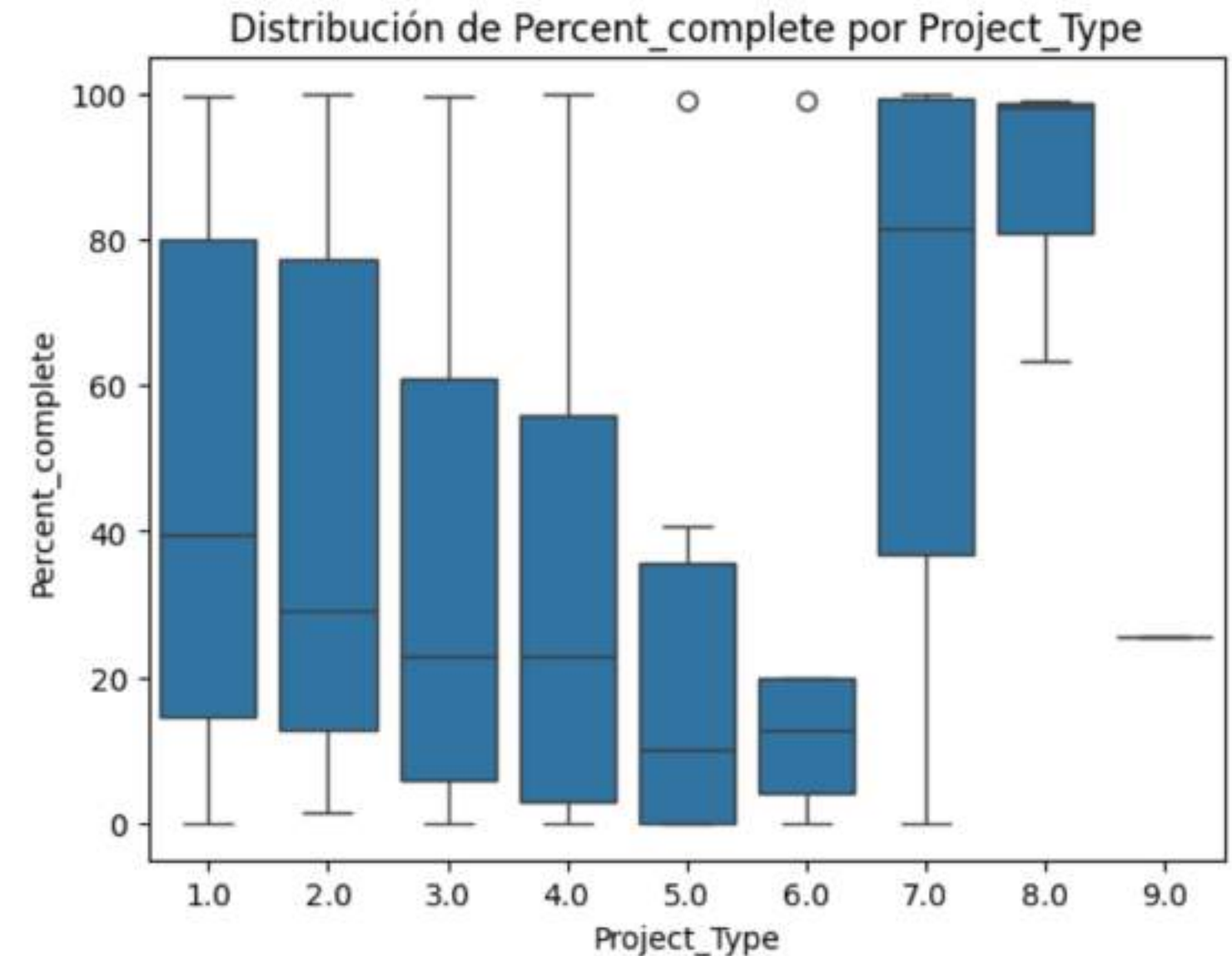


Distribución de Percent_complete por Project_Type

El porcentaje de avance depende del estado del proyecto, pero no del tipo de proyecto.

Los proyectos en progreso (Work in progress) presentan mayores valores de Percent_complete sin importar el tipo.

Fuente de variación	F	p-value (PR(>F))
State	9.20	0.00014
Project_Type	1.79	0.0792
State × Project_Type	0.96	0.497



Decisión:

Se rechaza H_0 únicamente para State, ya que $p < 0.05$.

Ni Project_Type ni la interacción entre ambos factores mostraron efectos significativos.

Distribución de Percent_complete por Project_Type

El avance de los proyectos depende principalmente del estado operativo (State).

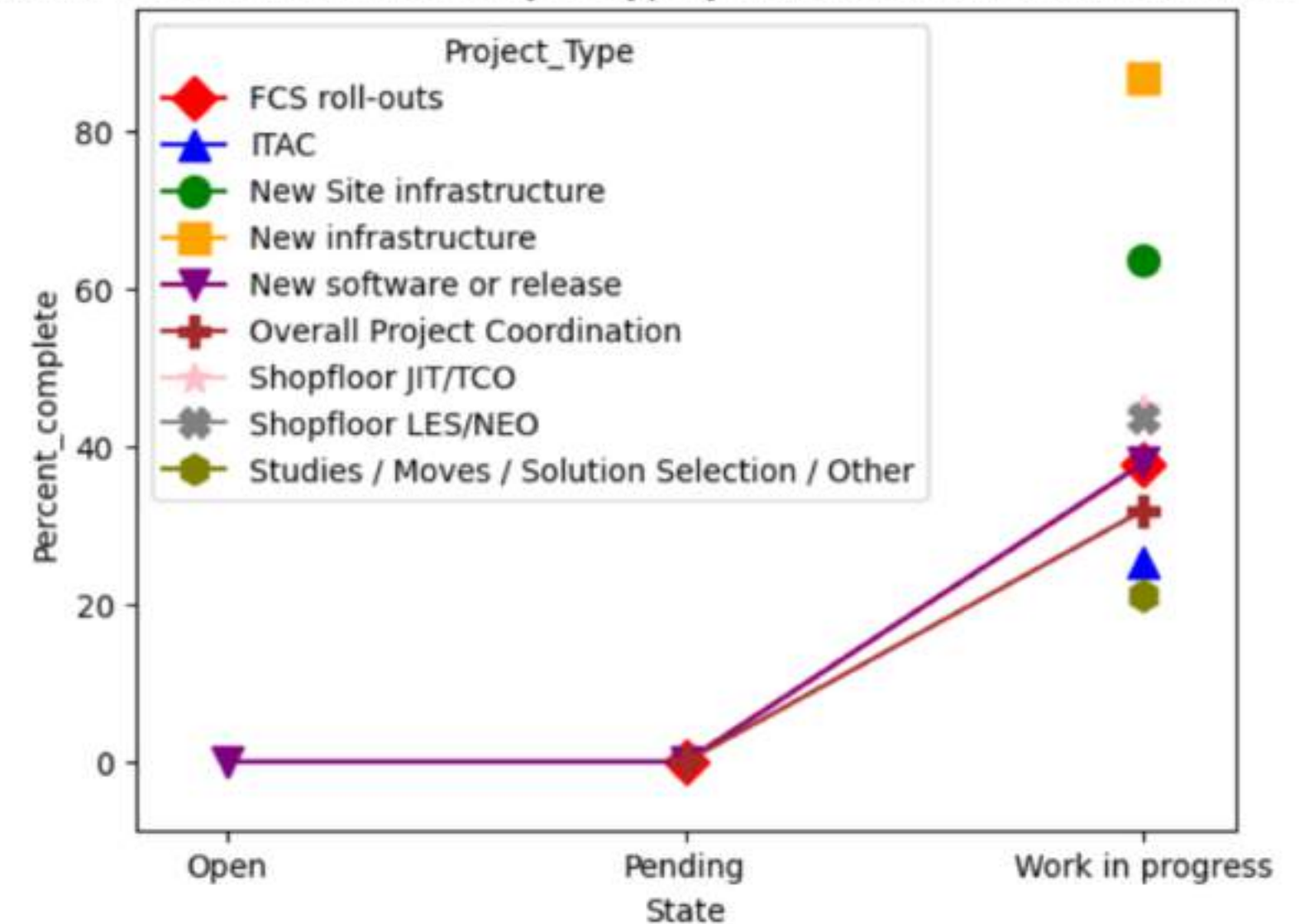
El tipo de proyecto no modifica este efecto, y la interacción entre ambos no es significativa.

Los proyectos "Work in progress" presentan mayor porcentaje de avance, sin importar el tipo.

No se encontró relación combinada entre tipo y estado.

Fuente de variación	F	p-value (PR(>F))
State	9.20	0.00014
Project_Type	1.79	0.0792
State × Project_Type	0.96	0.497

Interacción entre la columna Project Type y State en la columna Percent complete



Se rechaza H_0 solo para el factor State ($p < 0.05$).
No hay efecto significativo de Project_Type ni de la interacción entre

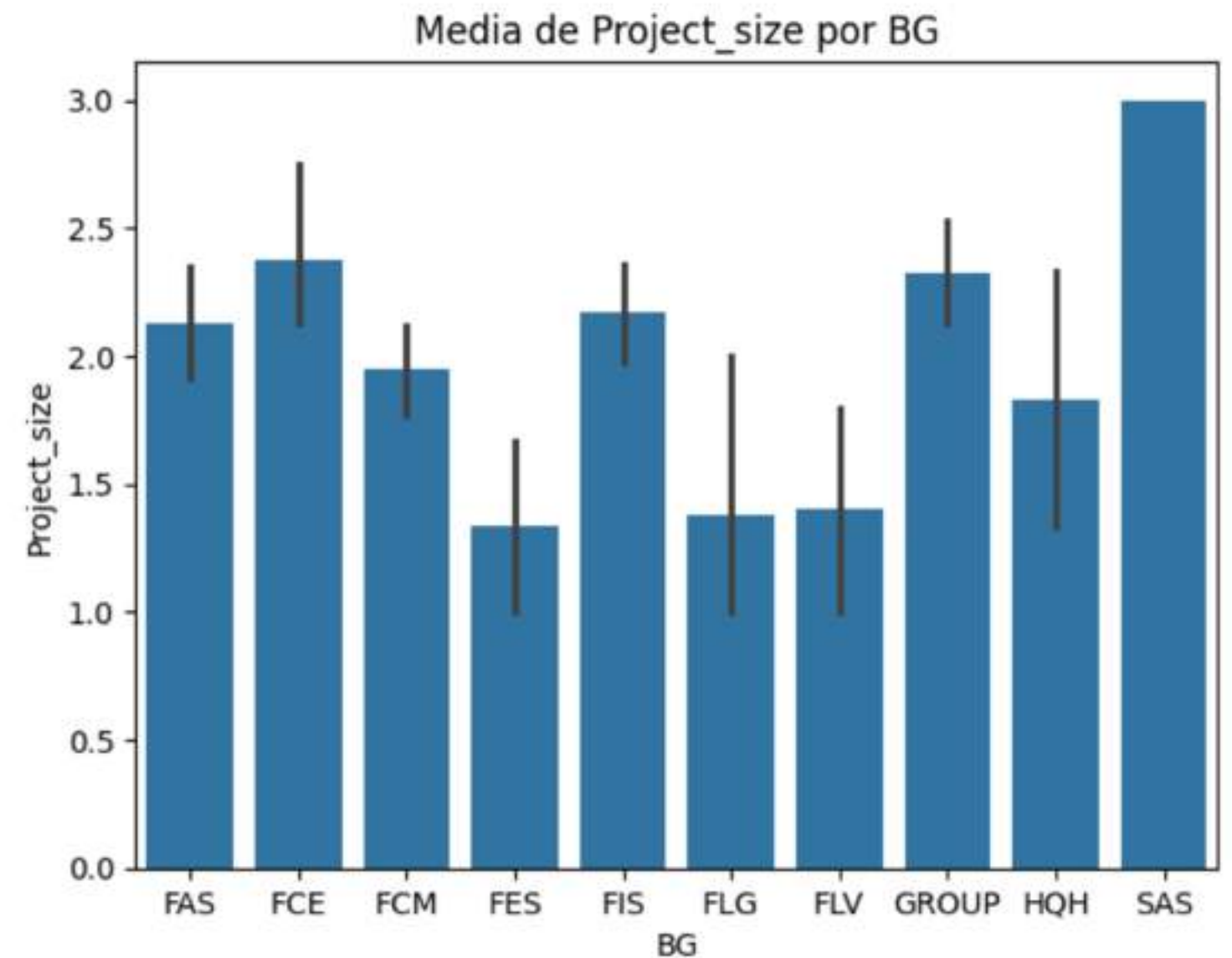
Project_size por BG

Los grupos SAS y GROUP presentan los mayores tamaños de proyecto, mientras que FES, FLG y FLV tienen los más pequeños.

Esto indica que el BG influye directamente en la magnitud de los proyectos.

El tamaño del proyecto depende significativamente del grupo de negocio (BG), lo que refleja diferentes niveles de alcance y recursos entre divisiones.

Fuente de variación	F	p-value (PR(>F))
BG	3.26	0.0009



Se rechaza H_0 , ya que $p < 0.05$.

Existen diferencias significativas en el tamaño promedio de los proyectos entre los distintos grupos de negocio

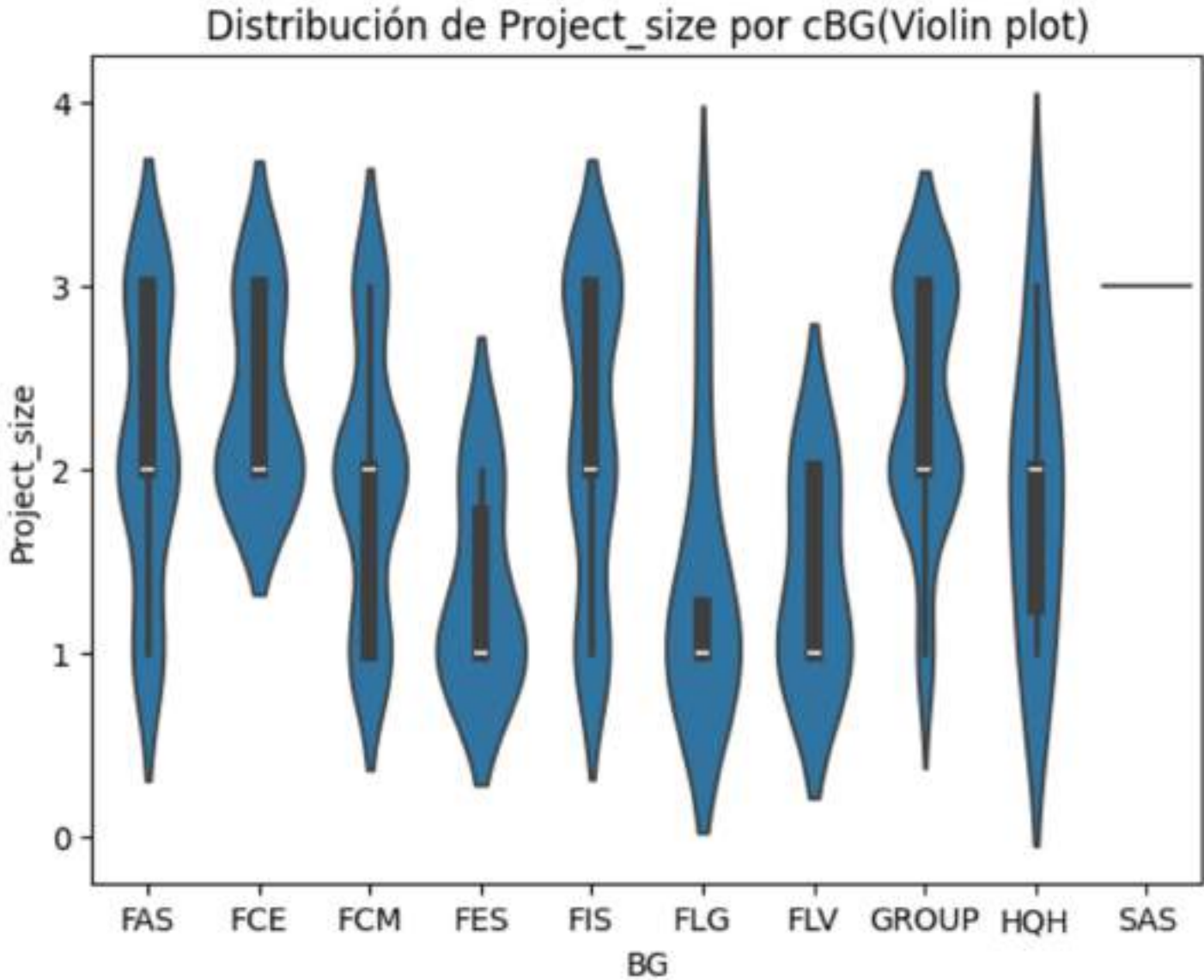
Distribución de Project_size por BG (Violin plot)

Los grupos SAS y GROUP concentran proyectos más grandes y con mayor variabilidad, mientras que FES, FLG y FLV muestran proyectos pequeños y más homogéneos.

Esto refleja diferencias claras en la escala de operación entre las divisiones.

El tamaño de los proyectos varía significativamente según el grupo de negocio (BG), lo que demuestra distintos niveles de complejidad y alcance dentro de la organización.

Fuente de variación	F	p-value (PR(>F))
BG	3.26	0.0009



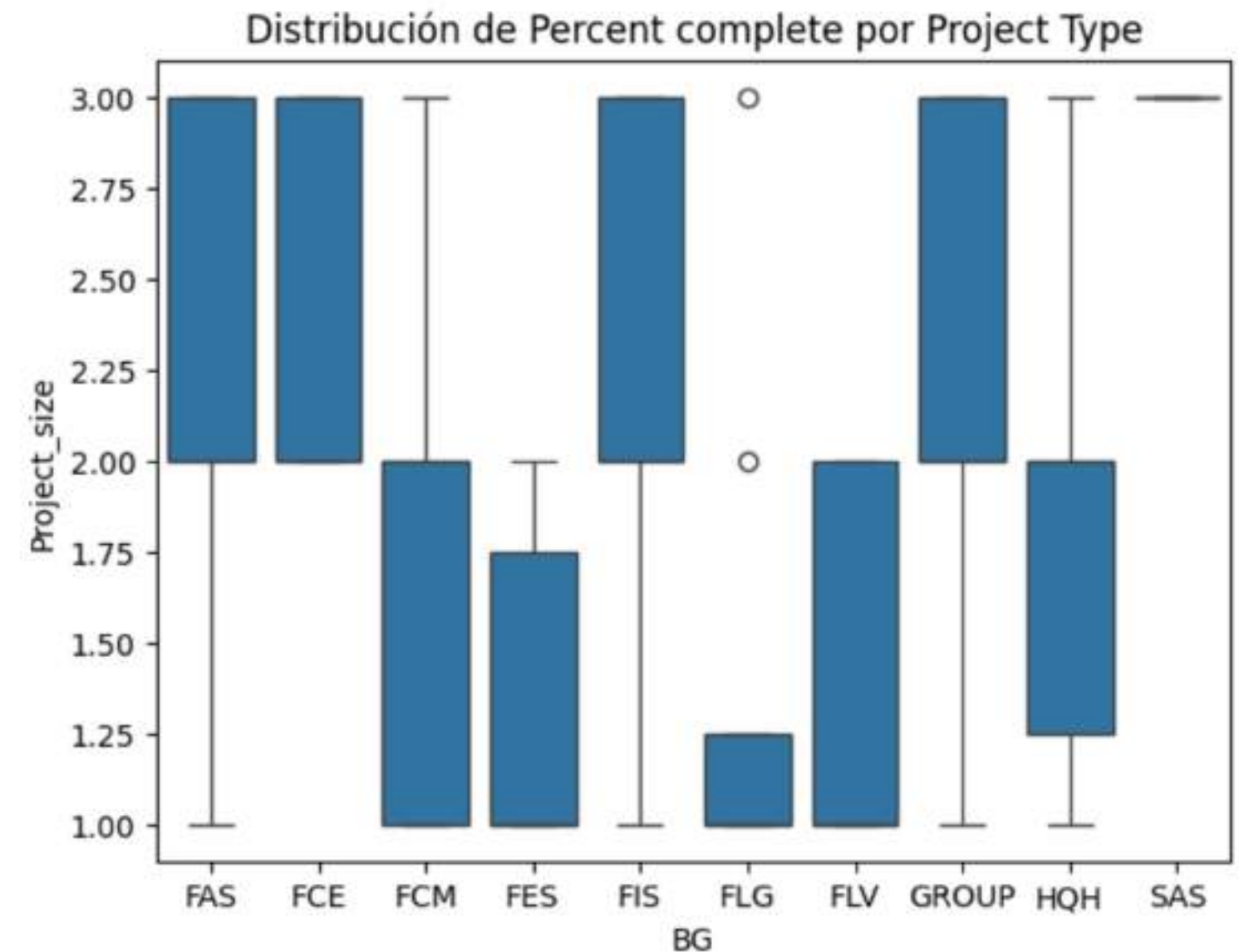
Se rechaza H_0 , ya que $p < 0.05$.
Existen diferencias significativas en el tamaño promedio de los proyectos entre los grupos de negocio.

Distribución de Percent complete por Project Type

Los grupos SAS, FAS y FCE concentran los proyectos más grandes,
mientras que FES, FLG y FLV presentan proyectos más pequeños y uniformes.

El grupo de negocio (BG) influye directamente en el tamaño del proyecto,
reflejando diferentes niveles de complejidad, recursos y alcance operativo entre divisiones.

Fuente de variación	F	p-value (PR(>F))
BG	3.26	0.0009



Se rechaza H_0 , ya que $p < 0.05$.

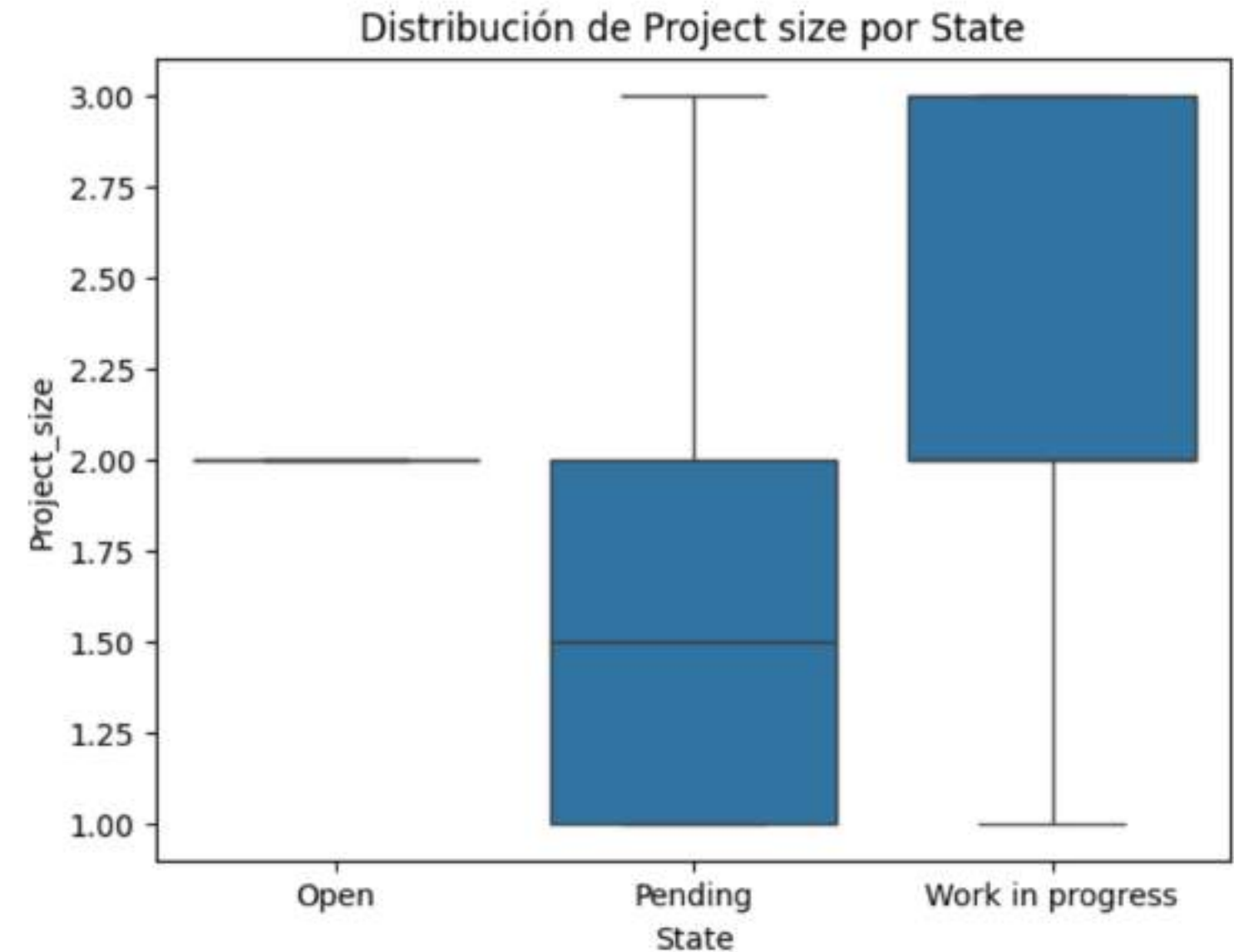
Existen diferencias significativas en el tamaño promedio de los proyectos entre los distintos grupos de negocio.

Distribución de Project_size por State

Los proyectos en "Work in progress" presentan los mayores tamaños y mayor variabilidad, mientras que los "Pending" y "Open" tienden a ser más pequeños y consistentes.

El estado operativo del proyecto influye en su tamaño. Los proyectos activos tienden a requerir mayor alcance y recursos, lo que los hace más grandes.

Fuente de variación	F	p-value (PR(>F))
BG	3.23	0.00010
State	1.19	0.3063
BG x State	1.09	0.3610



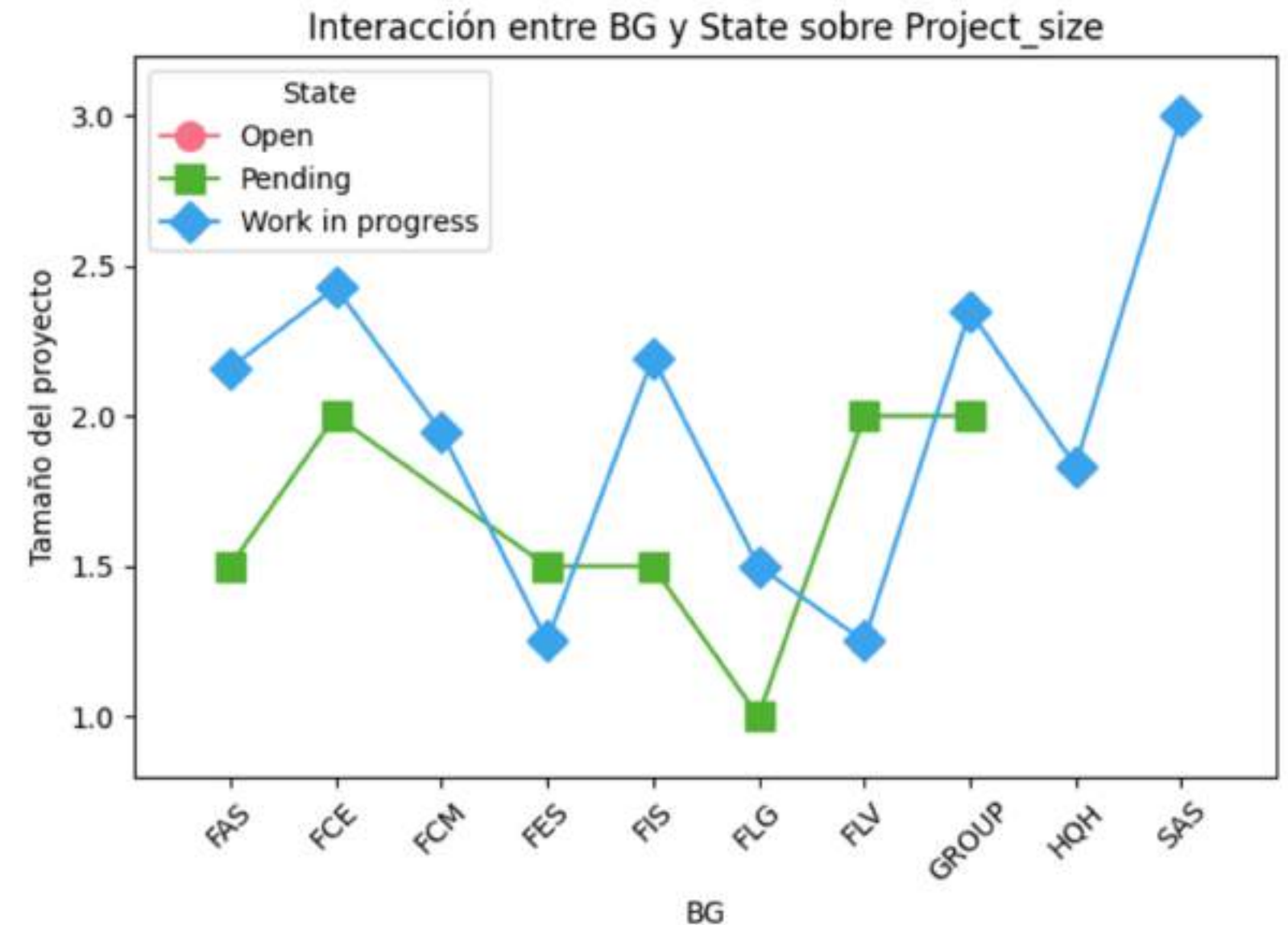
Se rechaza H_0 , ya que $p < 0.05$.
Existen diferencias significativas en el tamaño de los proyectos entre los distintos estados.

Interacción entre BG y State sobre Project_size

El grupo de negocio (BG) influye significativamente en el tamaño del proyecto, mientras que el estado del proyecto (State) no afecta el tamaño ni modifica el efecto del BG.

El tamaño de los proyectos depende principalmente del grupo de negocio, y esta relación se mantiene estable sin importar el estado operativo del proyecto.

Fuente de variación	F	p-value (PR(>F))
BG	3.23	0.00010
State	1.19	0.3063
BG x State	1.09	0.3610



Se rechaza H_0 únicamente para BG ($p < 0.05$).
Ni State ni la interacción BG \times State muestran diferencias significativas.

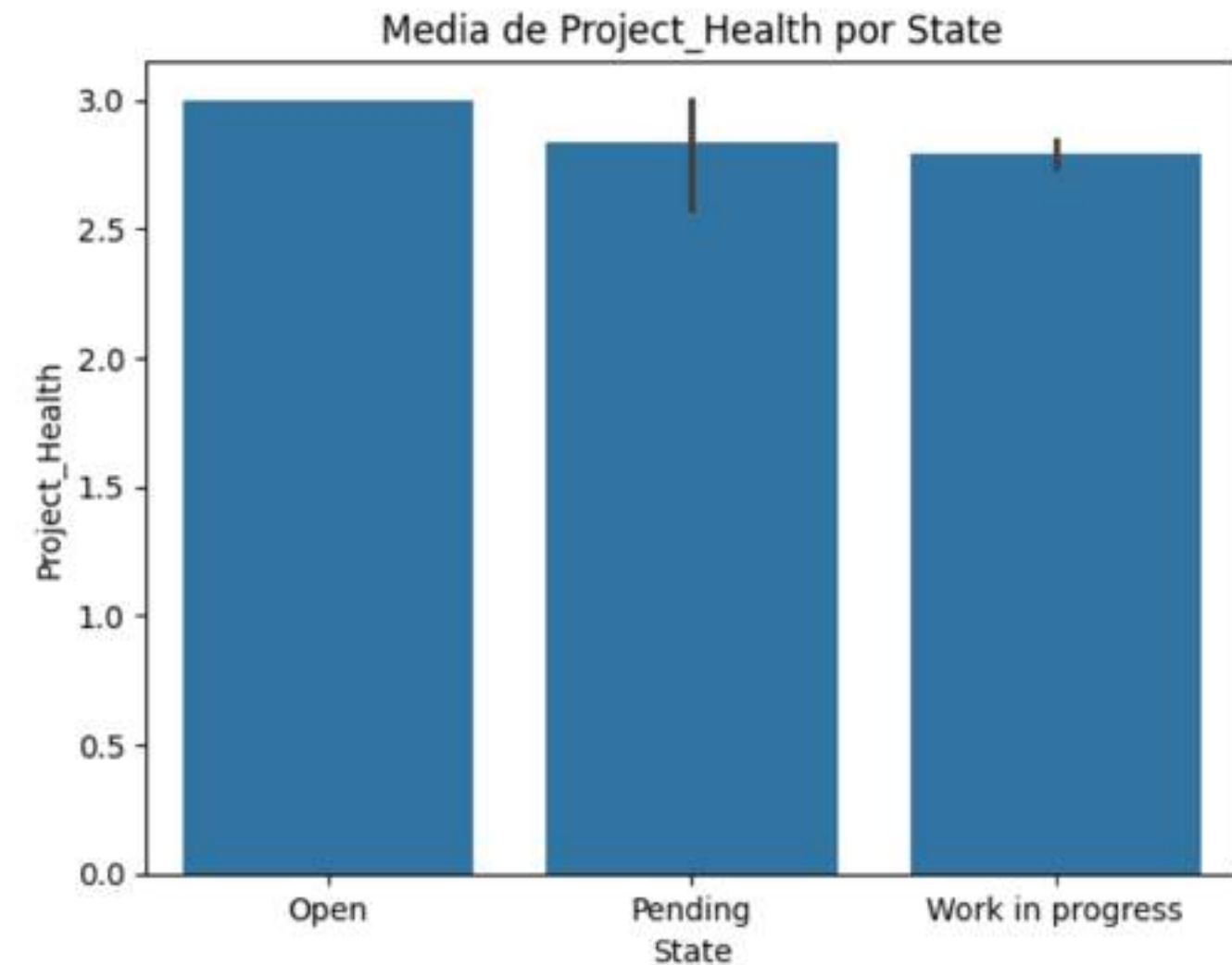
Media de Project_Health por State

La salud promedio de los proyectos se mantiene similar en todos los estados (Open, Pending y Work in progress).

Aunque los proyectos abiertos muestran una leve ventaja, la diferencia no es estadísticamente relevante.

El estado operativo no influye en la salud del proyecto; todos los proyectos mantienen niveles estables de Project_Health.

Fuente de variación	F	p-value (PR(>F))
State	1.18	0.3063



No se rechaza H_0 , ya que $p > 0.05$.

No existen diferencias significativas entre los estados del proyecto.

Distribución de Project_Health por State (Violin plot)

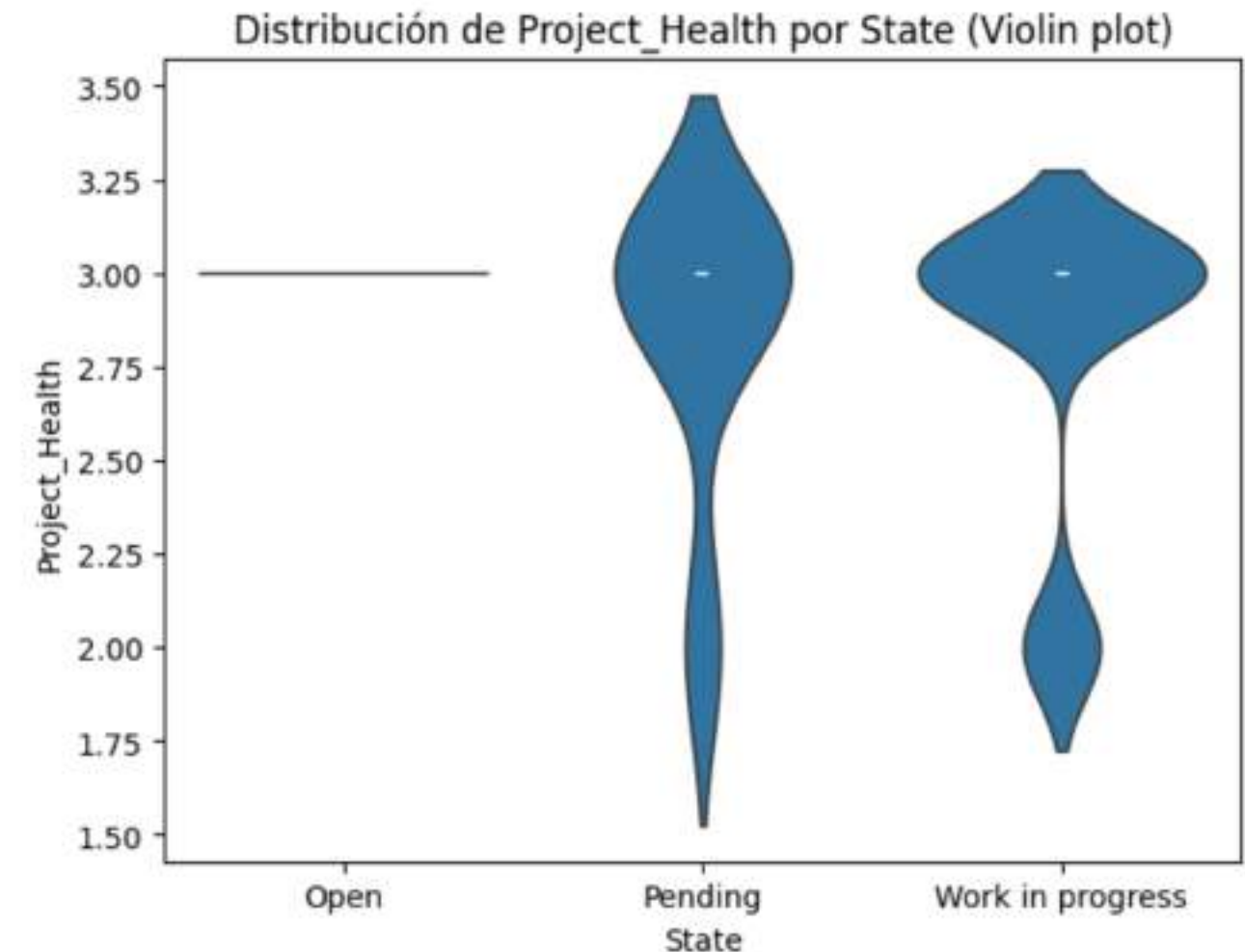
La salud promedio de los proyectos es similar en los tres estados (Open, Pending, Work in progress).

Aunque los proyectos abiertos parecen tener una leve ventaja, la variabilidad es baja y no se considera estadísticamente relevante.

El estado del proyecto (State) no afecta de manera significativa su salud general (Project_Health).

Esto sugiere que la gestión y estabilidad del proyecto se mantiene constante sin importar la etapa operativa.

Fuente de variación	F	p-value (PR(>F))
State	1.18	0.3063



No se rechaza H_0 , ya que $p > 0.05$.

No existen diferencias significativas en la salud de los proyectos según su estado.

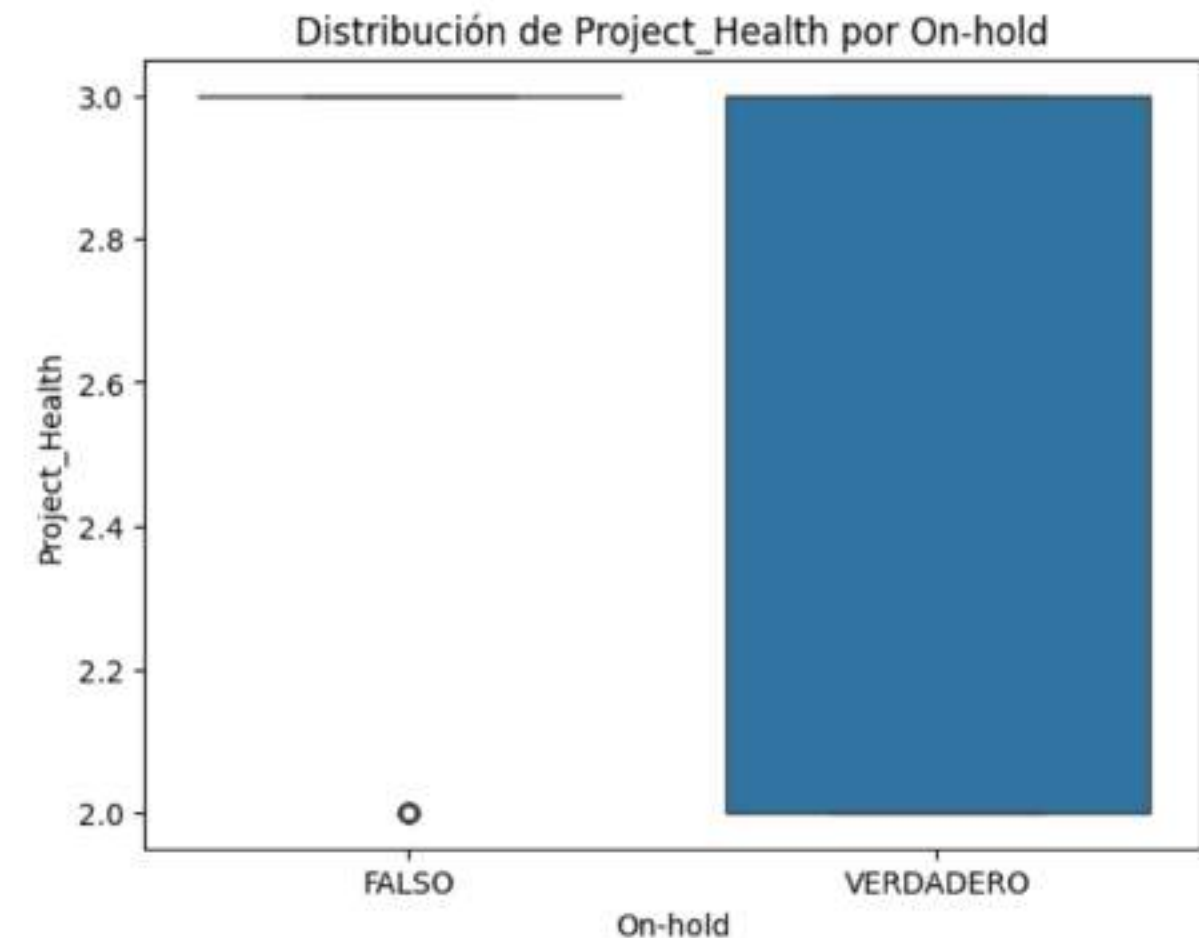
Distribución de Project_Health por On-hold

Los proyectos en pausa (On-hold = VERDADERO) muestran una salud promedio más baja y menor estabilidad, mientras que los proyectos activos (On-hold = FALSO) tienden a mantener una salud más alta y constante

El estado de pausa (On-hold) influye significativamente en la salud del proyecto (Project_Health).

Los proyectos pausados reflejan mayor riesgo o deterioro en su gestión

Fuente de variación	F	p-value (PR(>F))
State	0.20	0.8158
On-hold	10.39	0.0014
State × On-hold	1.03	0.3584



Se rechaza H_0 , ya que $p < 0.05$.

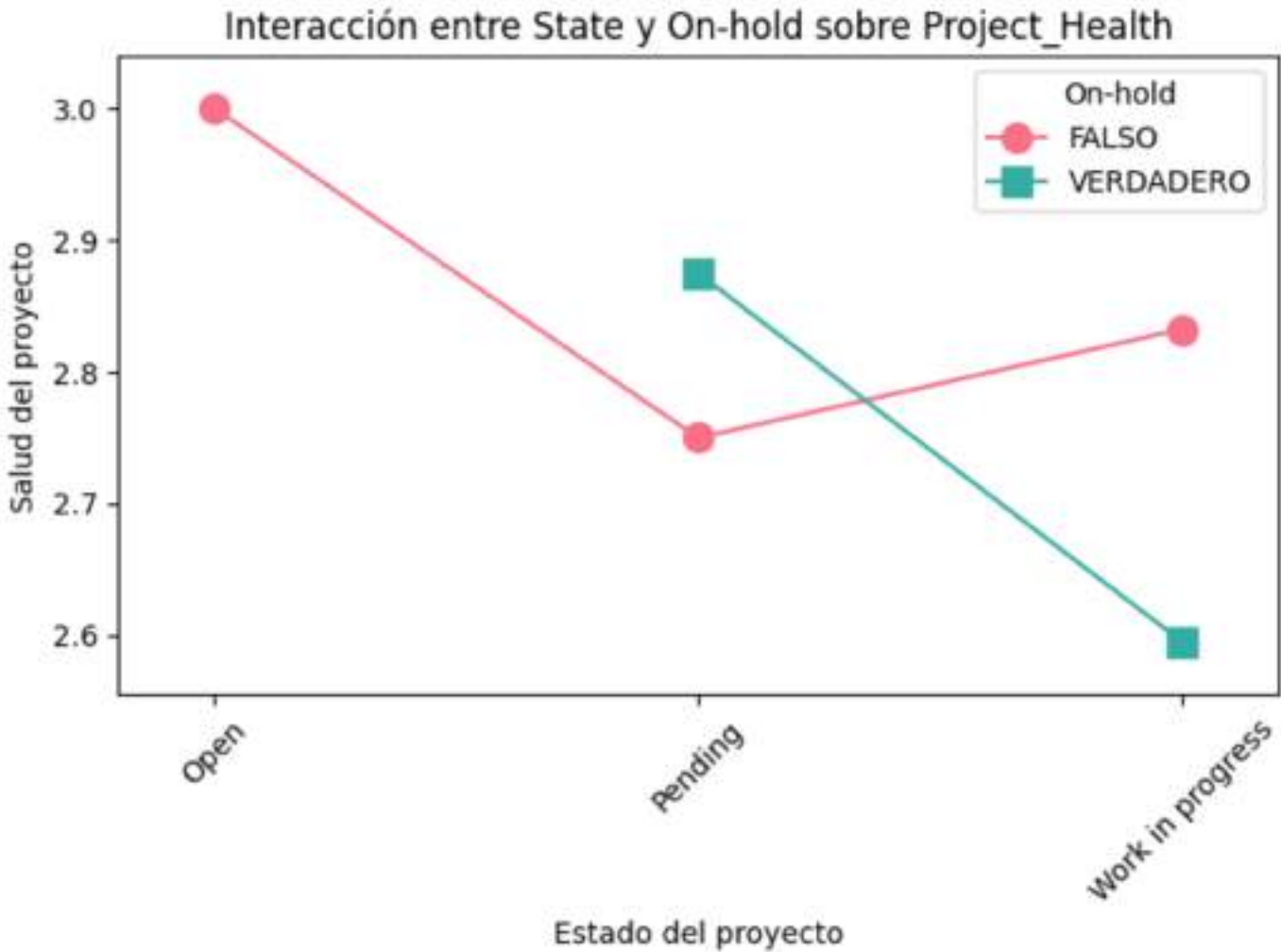
Existen diferencias significativas en la salud de los proyectos dependiendo de si están pausados o no.

Interacción entre State y On-hold sobre Project_Health

El efecto más relevante proviene del estado de pausa (On-hold), ya que los proyectos pausados muestran una menor salud promedio, especialmente en la etapa Work in progress. No se observa interacción significativa: el patrón entre estados es similar tanto para proyectos activos como pausados.

El factor On-hold tiene un impacto directo y negativo en la salud del proyecto, mientras que el estado operativo no modifica significativamente esta tendencia.

Fuente de variación	F	p-value (PR(>F))
State	0.20	0.8158
On-hold	10.39	0.0014
State × On-hold	1.03	0.3584



Se rechaza H_0 únicamente para On-hold, ya que $p < 0.05$.
No hay evidencia de efectos significativos para State ni para la interacción entre ambos factores.

Media de Project_Type por Project_size

Interpretación:

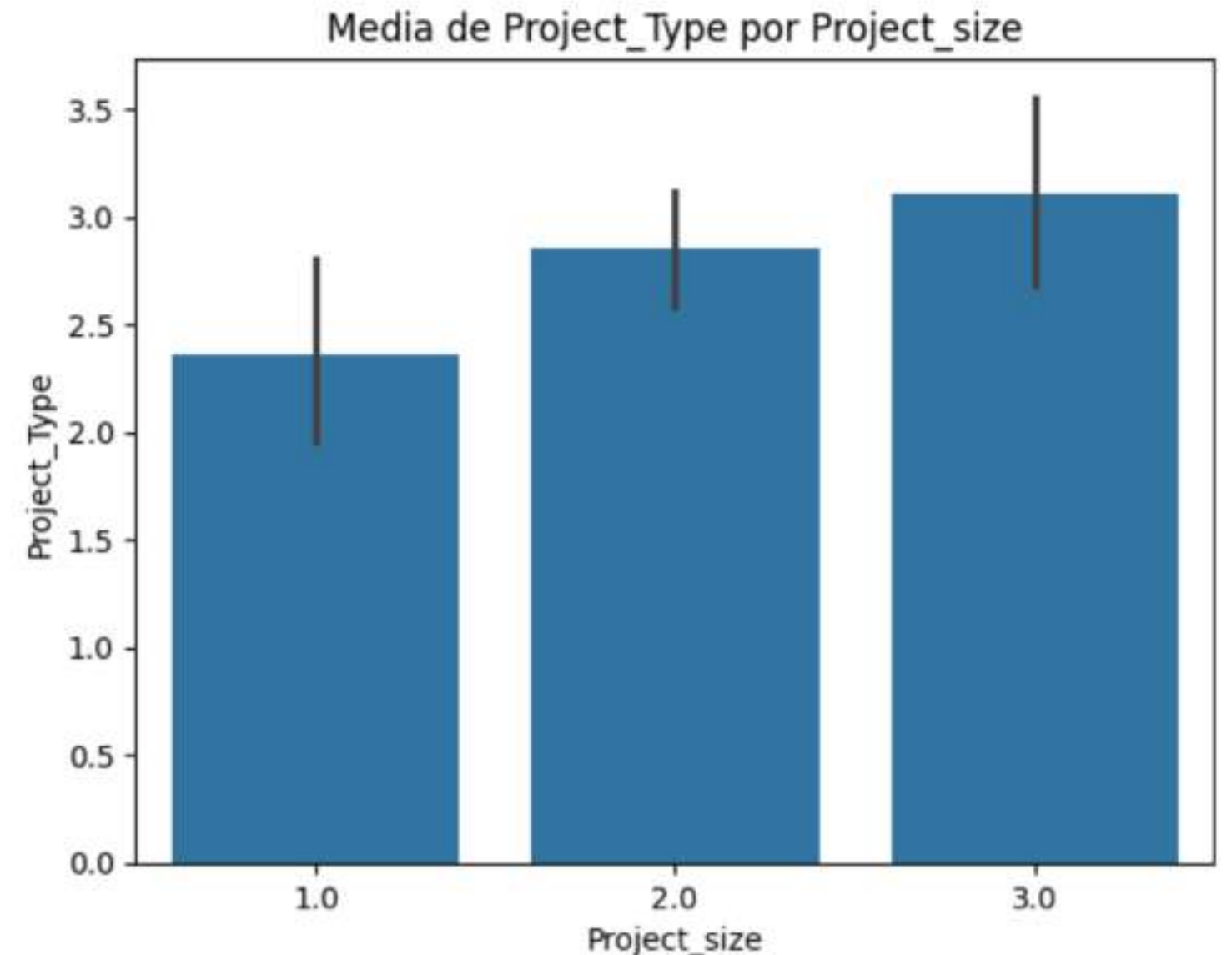
El gráfico muestra que a medida que aumenta el tamaño del proyecto, también crece el tipo de proyecto asociado.

- Los proyectos pequeños (1) corresponden principalmente a tipos más simples.
- Los medianos (2) y grandes (3) tienden a estar vinculados con proyectos más complejos y estratégicos.

Resultado ANOVA:

$p = 0.03 \rightarrow$ existen diferencias significativas entre los tamaños.

Fuente de variación	F	p-value (PR(>F))
Project size	3.53	0.030



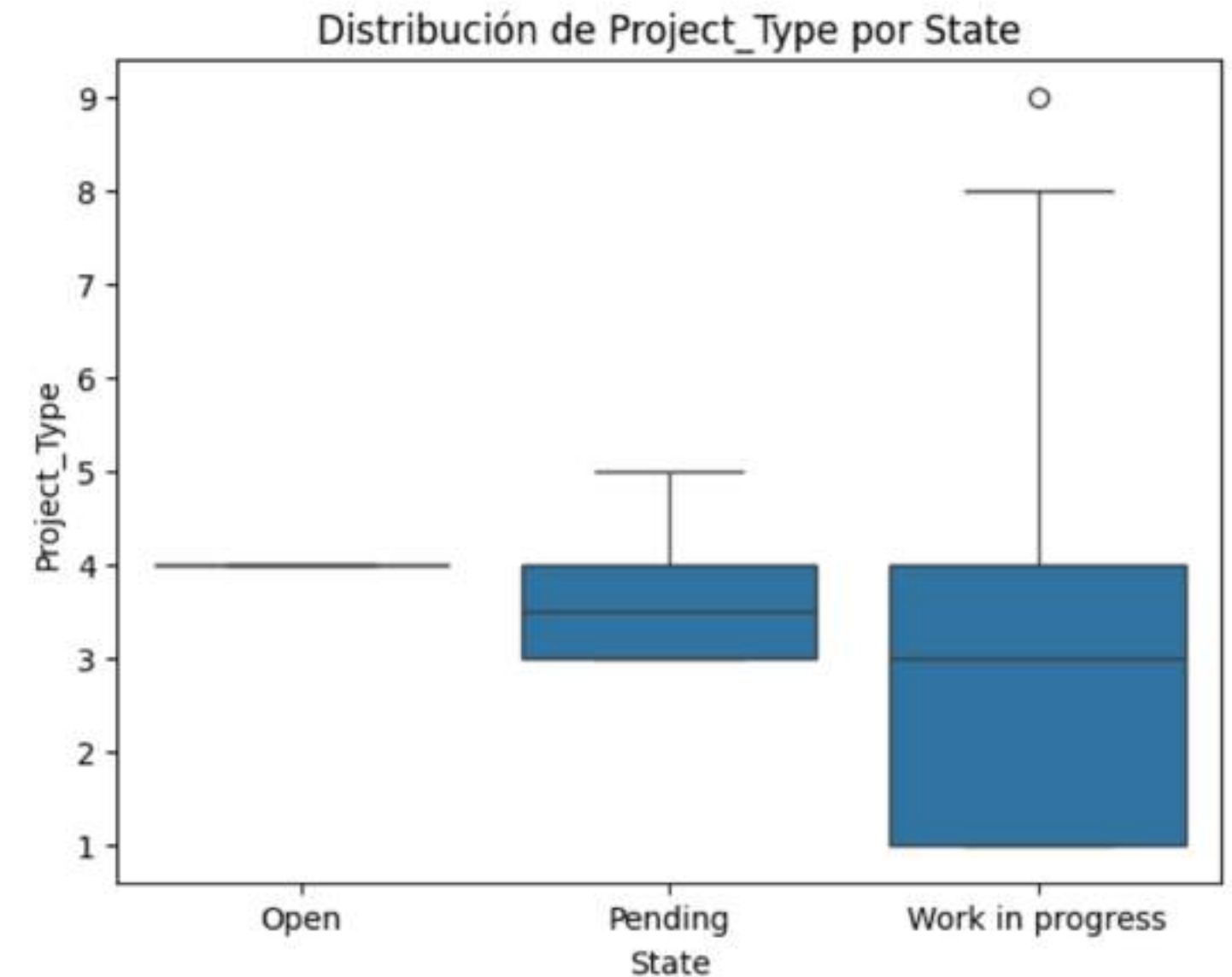
Distribución de Project_Type por State

Interpretación:

El gráfico muestra cómo varía el tipo de proyecto (Project_Type) según el estado (State).

- Los proyectos "Work in progress" presentan mayor variabilidad, incluyendo tipos más complejos.
- En "Pending", los tipos se concentran en niveles medios.
- Los "Open" muestran tipos más homogéneos

Fuente de variación	F	p-value (PR(>F))
Project size	3.55	0.0309
State	2.31	0.101
Project size x State	0.44	0.77



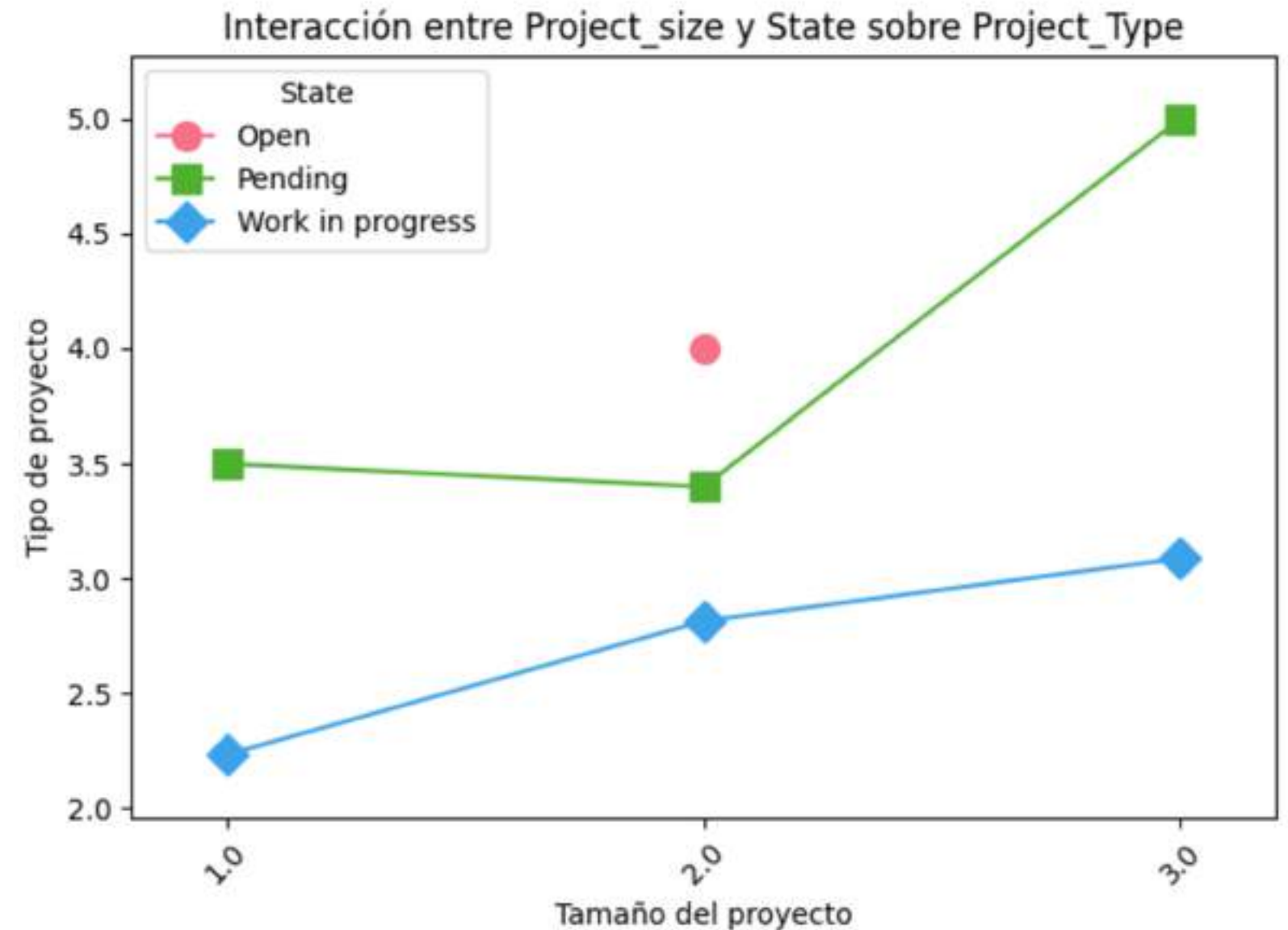
Interacción entre Project size y State sobre Project_Type

Interpretación:

Se observa una interacción moderada entre el tamaño y el estado del proyecto:

- En proyectos pequeños, el tipo se mantiene bajo sin importar el estado.
- En proyectos medianos y grandes, los proyectos pendientes tienden a ser de mayor tipo o complejidad.
- Los en progreso aumentan de forma gradual con el tamaño.

Fuente de variación	F	p-value (PR(>F))
Project size	3.55	0.0309
State	2.31	0.101
Project size x State	0.44	0.77



ANÁLISIS DE INSIGHTS

STATE (Estado del Proyecto) es el factor más influyente

Las interacciones son cruciales

Los proyectos en pausa necesitan métricas especiales



**¡MUCHAS
GRACIAS!**