

Reporte de Análisis Estadístico

Equipo 4

Introducción

El presente reporte tiene como objetivo analizar la variación en la abundancia de diatomeas (algas microscópicas) utilizando métodos estadísticos como ANOVA (Análisis de Varianza) y MANOVA (Análisis de Varianza Multivariado).

El análisis se basó en el archivo “01_DiatomInventories_GTstudentproject_B.csv”, el cual contiene información sobre distintas especies de diatomeas, sitios de muestreo y fechas de recolección. Se filtraron los 15 valores más frecuentes de cada variable categórica para facilitar la comparación y reducir la dispersión del conjunto de datos.

Metodología

1. Limpieza y filtrado de datos:

Se seleccionaron los 15 valores más frecuentes de las variables categóricas:

- *TaxonName*
- *TaxonCode*
- *SamplingOperations_code*
- *CodeSite_SamplingOperations*
- *Date_SamplingOperation*

2. Análisis descriptivo y visualización:

Se realizaron boxplots, gráficas de barras y violin plots para observar la distribución de las variables dependientes:

- *Abundance_nbccl*
- *TotalAbundance_SamplingOperation*
- *Abundance_pm*

3. Análisis de varianza (ANOVA):

Se aplicó ANOVA unifactorial para determinar si las diferencias entre grupos (por taxón, sitio o fecha) eran estadísticamente significativas.

4. ANOVA multifactorial (MANOVA):

Se emplearon combinaciones de variables independientes (como *TaxonName* y *SamplingOperations_code*) para analizar la influencia conjunta de factores biológicos y de muestreo sobre la abundancia.

Resultados

Variable: Abundance_nbcell

- **ANOVA unifactorial:**
 - *TaxonName* mostró un efecto significativo sobre la abundancia celular ($p = 0.000008 < 0.05$).
 - Esto indica diferencias importantes en la abundancia de células entre los distintos taxones.
- **MANOVA:**
 - *CodeSite_SamplingOperations* y *Date_SamplingOperation* no tuvieron efectos significativos ($p > 0.95$).
 - La abundancia celular no varía significativamente según el sitio ni la fecha del muestreo.

Interpretación:

Las diferencias en *Abundance_nbcell* se deben principalmente a características biológicas del taxón y no a factores espaciales o temporales.

Variable: TotalAbundance_SamplingOperation

- **ANOVA unifactorial:**
 - *TaxonCode* no mostró efecto significativo sobre la abundancia total ($p = 1.0$). No existen diferencias relevantes entre los códigos de taxón.
- **MANOVA:**
 - *TaxonName* y *SamplingOperations_code* sí tuvieron efectos significativos ($p \approx 0.0$).
 - Esto indica que tanto la especie como el tipo de operación de muestreo influyen en la abundancia total.

Interpretación:

Aunque el código taxonómico no refleja diferencias por sí solo, la combinación de nombre del taxón y el código de muestreo sí impacta significativamente la abundancia total registrada.

Variable: Abundance_pm

- **ANOVA unifactorial:**

- *SamplingOperations_code* no mostró efecto significativo ($p = 0.96$).
- No se observan diferencias estadísticas en la abundancia promedio entre operaciones de muestreo.
- **MANOVA:**
 - *Date_SamplingOperation* y *CodeSite_SamplingOperations* tampoco mostraron efectos significativos ($p > 0.95$).
 - La *abundance_pm* no varía significativamente con el sitio ni con la fecha del muestreo.

Interpretación:

Las fluctuaciones en *Abundance_pm* no se asocian con la ubicación o el momento del muestreo, lo cual sugiere estabilidad temporal y espacial en esta variable.

Discusión

Los análisis realizados evidencian que las variables biológicas (como el *TaxonName*) tienen una mayor influencia sobre la abundancia de diatomeas que las variables espaciales o temporales (sitio o fecha de muestreo).

Este resultado es coherente con la idea de que diferentes especies poseen distintas tasas de crecimiento y distribución poblacional, mientras que las condiciones ambientales en los sitios analizados pueden haber sido similares, reduciendo su efecto sobre las abundancias observadas.

Asimismo, el análisis multifactorial permitió identificar que la interacción entre factores biológicos y operativos (como el tipo de muestreo) puede generar diferencias significativas en ciertos indicadores de abundancia.

Conclusiones

- El taxón es el principal factor determinante de la abundancia celular y total en el conjunto de datos analizado.
- No se detectaron efectos significativos del sitio ni de la fecha de muestreo en las variables de abundancia.
- La abundancia total se ve influida por la combinación de *TaxonName* y *SamplingOperations_code*, lo cual sugiere una interacción entre características biológicas y métodos de muestreo.

- En general, los factores biológicos tienen mayor impacto en la abundancia de diatomeas que los factores ambientales o operativos.

Recomendaciones

- Ampliar el tamaño de muestra o incluir más sitios para fortalecer la validez de los resultados.
- Realizar análisis multivariados adicionales (por ejemplo, PCA o regresión múltiple) para identificar posibles correlaciones entre variables ambientales y biológicas.
- Complementar el estudio con variables fisicoquímicas del agua (pH, temperatura, nutrientes) que podrían explicar parte de la variabilidad no significativa detectada.