

Análisis Exploratorio y Estadístico

Mortalidad en Salmo salar por bloom de algas y OD

Felipe Tucca Díaz

INTESAL

2022-06-30

Estructura del trabajo exploratorio y estadístico

1) Introducción

- Descripción de la problemática.

2) Análisis exploratorio de los datos

- Histograma biomasa muerta (toneladas) por causa (Bloom algas/Oxígeno disuelto).
- Boxplot biomasa muerta por causa entre el 2011 al 2022.

3) Análisis estadístico de los datos

- Modelos lineales simples y múltiple.
- Comparación de modelos usando RSS-AIC.

3) Conclusiones

Introducción

Descripción de la problemática

- Base de datos presenta registros de mortalidad por causa **bloom de algas y disminución de oxígeno disuelto (OD)**.
- 23 centros de cultivos reportaron la causal de mortalidad en salmones para un barrio de la Región de Los Lagos.
- El salmón del Atlántico (*Salmo salar*) es la especie más cultivada en el barrio.
- Los registros corresponden a mortalidades entre los años 2011 e inicio del 2022 (Total de registros= 1224).
- Las variables de estudio fueron causa mortalidad, peso (g) del salmón, años, mes, semana de registro y la identificación de cada centro de cultivo que operó entre el 2011 al 2022.

Objetivos del estudio

- Evaluar la causa de mortalidad por bloom de algas y OD sobre la especie *Salmo salar* para un barrio del sur de Chile entre los años 2011 a inicios del 2022.
- Generar un modelo lineal que mejor ajuste la predicción de mortalidad en la biomasa de salmones.

Análisis exploratorio de los datos

- Existen datos desbalanceados por causa de muertes en salmonidos: Causas bloom de algas ($n=360$) y OD ($n=864$).
- **Año 2021** presentó una mayor biomasa muerta para los últimos 10 años.
- No existe correlación significativa ($p<0.05$) entre las causas de muerte por bloom de algas y disminución de OD en el barrio.
- **Ha existido una mayor biomasa muerta (ton) en el barrio a causa del bloom de algas.**

HISTOGRAMA

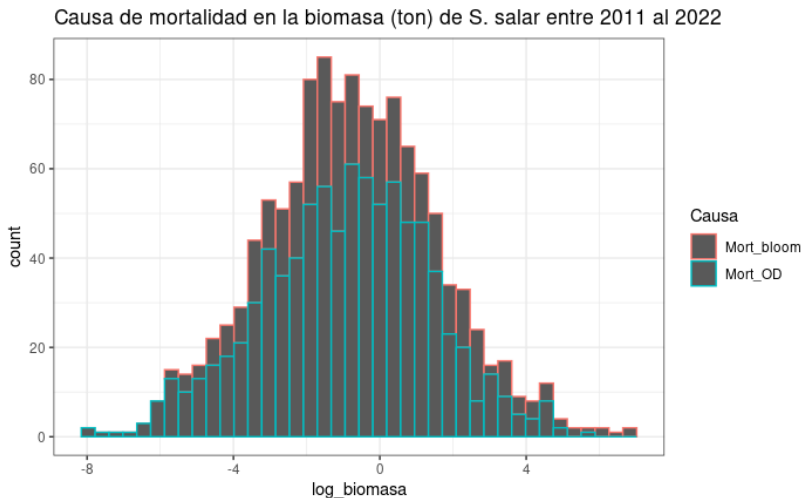


Figure 1: Histograma biomasa muerta (toneladas) por causa

Boxplot: Datos faltantes y datos atípicos

- Boxplot consideró la causa de muerte sobre la biomasa de peces entre el 2011 al 2022.
- Existen datos faltantes principalmente para la causa bloom de algas.
- Valores atípicos se presentaron para los dos casos de mortalidad en salmones.

BOXPLOT

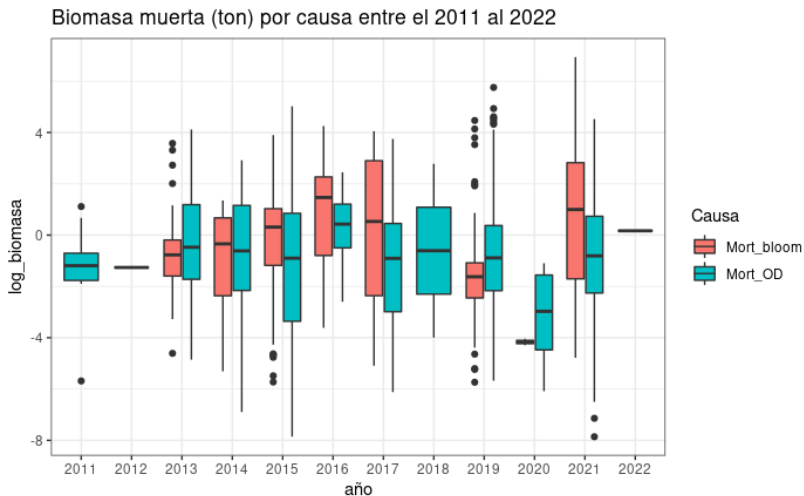


Figure 2: Boxplot biomasa muerta por año y causa

Biomasa muerta en relación a la semana de registro

- Se evidencia la ocurrencia de un evento temporal puntual que generó una alta mortalidad en la biomasa de salmones.
- Año 2021 presentó la mayor mortalidad registrada históricamente en este barrio (Log biomasa muerta > 5).
- La mortalidad por **bloom de algas** para el años 2021 generó *valores atípicos* de mortalidad en el barrio.

Relación biomasa muerta y las semanas que se reportó mortalidad

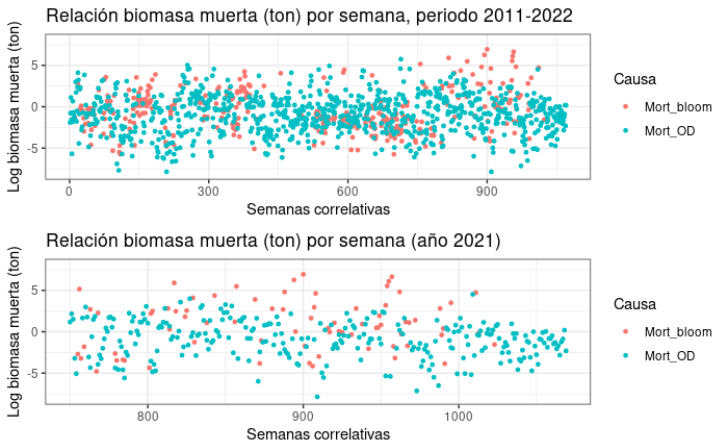


Figure 3: Biomasa muerta por semana y causa

Análisis exploratorio de los datos: Mortalidad bloom vs OD

- La mayor biomasa muerta es por causa de bloom de algas, alcanzando 16 toneledas en los ultimos 10 años.
- La mortalidad por OD alcanzó las 4.1 toneladas.
- El peso promedio de los salmones muertos fue de 3.1 kilogramos (+/- 1.5 kg.)

Tabla resumen de la biomasa muerta por causa

Table 1: Resumen de la biomasa muerta (toneladas) para la especie *S. salar* por causa entre los años 2011 al 2021

| Causa | N | Promedio | DE | Mediana |
|------------|-----|-----------|----------|-----------|
| Mort_bloom | 360 | 16.184949 | 82.38176 | 0.4478442 |
| Mort_OD | 864 | 4.071008 | 16.81336 | 0.4349179 |

Análisis estadístico de los datos: Modelo lineal simple

- Se realizó un **modelo de regresión lineal simple** con los factores centro, semanas y años.
- Los modelos fueron estadísticamente significativos ($p < 0.05$), pero con un bajo ajuste o R^2 ajustado menor al 7%.

Hipótesis modelo lineal simple

- Basado en estos modelos de regresión simple se rechazó hipótesis nula que postuló:

Hipótesis nula (H0): Existe similitud en la biomasa muerta entre centros/semanas/años.

- se aceptó H1 con un $p < 0.05$:

Hipótesis alternativa (H1): No existe similitud en la biomasa muerta entre centros/semanas/años.

Hipótesis modelo lineal múltiple

- Para el **modelo de regresión múltiple** se postularon las siguientes hipótesis:

H0:

$$\beta_j = 0; j = 1, 2, \dots, k$$

H1:

$$\beta_j \neq 0; j = 1, 2, \dots, k$$

- El modelo cumplió con los tres supuestos: linealidad, homogeneidad de varianza y normalidad.

Análisis estadístico de los datos: Ajuste modelo lineal múltiple

- La modelación integró los factores causa, centro, año, mes y la interacción entre causa y año.
- El modelo nos entrega como resultado coeficientes distintos de cero, por lo tanto, se rechaza la H_0 (valores p menores al 5%).
- El valor R^2 ajustado de esta modelación múltiple correspondió a un 23%.

Análisis de varianza (ANOVA)

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|------|------------|-----------|-----------|-----------|
| Causa | 1 | 63.15504 | 63.155039 | 14.624314 | 0.0001381 |
| año | 11 | 248.97448 | 22.634044 | 5.241187 | 0.0000000 |
| centro_id | 22 | 615.45852 | 27.975387 | 6.478040 | 0.0000000 |
| mes | 11 | 665.02969 | 60.457244 | 13.999607 | 0.0000000 |
| Causa:año | 7 | 175.41762 | 25.059660 | 5.802868 | 0.0000012 |
| Residuals | 1171 | 5056.95851 | 4.318496 | NA | NA |

Comparación de modelos por RSS y AIC

- Fueron usados criterios de residuales (RSS) y Akaike (AIC)
- Ambos criterios sugieren que el modelo lineal múltiple presenta mejor predicción y ajuste (23%).

| Res.Df | RSS | Df | Sum of Sq | F | Pr(>F) |
|--------|----------|-----|------------|-----------|-----------|
| 1201 | 6252.877 | NA | NA | NA | NA |
| 1212 | 6270.714 | -11 | -17.83663 | 0.3754807 | 0.9656779 |
| 1212 | 6550.265 | 0 | -279.55122 | NA | NA |
| 1171 | 5056.959 | 41 | 1493.30669 | 8.4339818 | 0.0000000 |

| | df | AIC |
|-----------------------|----|----------|
| modelo1_anova1_centro | 24 | 5517.805 |
| modelo2_anova2_mes | 13 | 5499.291 |
| modelo3_anova3_año | 13 | 5552.676 |
| lm.aov_biomasa | 54 | 5317.978 |

Interpretación y conclusiones del trabajo

- Análisis exploratorio muestra mayor mortalidad de la biomasa de peces en el barrio por bloom de algas.
- Mortalidades debido a bajas de OD presentaron mayor frecuencia de registro.
- Se realizó ANOVA con un vía de criterio de clasificación para los factores centro de cultivo, semanas y años con ajustes menores al 7%.
- Modelo lineal múltiple agrupó todas los factores mostrando significativamente un mejor ajuste de la predicción para la variable biomasa muerta.
- Análisis comparativo por RSS y AIC determinaron que la **regresión lineal múltiple representa una mejor predicción**