

# Mortalidad por Bajas de O<sub>2</sub> en función a la concentración de O<sub>2</sub>

Diplomado en Análisis de datos con R para la Acuicultura.

Gabriela Salas Serqueira

Pontificia Universidad Católica de Valparaíso

2022-06-30

# Introducción

**Descripción del problema:** La mortalidad en salmones de cultivo es uno de los problemas que enfrenta la industria salmonera desde sus inicios, en la actualidad, los cambios en el clima ha afectado particularmente a esta industria, en cuanto a variaciones en la temperatura del agua, horas luz, salinidad y variaciones de oxígeno disuelto entre otros, en este estudio intentaremos dilucidar si existen una tendencia según la ubicación de las unidades de cultivo dentro de la X Región y la mortalidad por bajas de oxígeno

**Descripción de las Variables de estudios:** Las variables que se usarán son los barrios donde están ubicados los centros dentro de la X Región (Agrupación de Centros), Mortalidad total (Total\_Mortalidad) y Mortalidad por bajas de Oxígeno (Mortalidad\_O2)

**Numero de observaciones:** 44

# Análisis Exploratorio

Tabla resumen de los datos a analizar.

Table 1: Estimadores puntuales

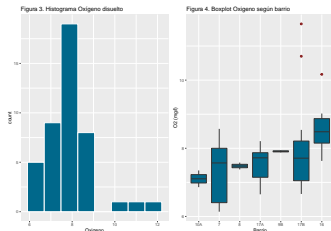
Barrio	n	mean(Mort_O2)	median(Mort_O2)	Variance
10A	2	5882.500	5882.5	69207612
16	8	859.000	52.0	3169801
17A	9	19780.222	25256.0	268998519
17B	14	12380.714	6558.0	270050518
7	7	2631.857	1202.0	18012292
8	2	33462.500	33462.5	188354640

Se observa que las observaciones no se encuentran balanceados, ya que los N muestrales en los distintos barrios son muy distintos entre ellos.

Ejemplo: Barrio 17B 14 muestras mientras que el barrio 8 sólo presenta 2 muestras.

# Exploratorio de datos

Mediante gráficos de histogramas revisamos la distribución de las variables de estudio



En los gráficos podemos observar que los datos de Oxígeno disuelto presenta una distribución normal, con algunos datos atípicos, la media de la concentración de oxígeno para el set de datos se encuentra al rededor de los 8 mg/l, existen mediciones fuera del rango normal, por sobre los 10mg/l. Es importante entender que estos datos representa los promedio de las mediciones de distintos ciclos productivos en los distintos barrios que se realizó el muestreo.

# Hipotesis prueba de correlación

Se realiza la correlación de Pearson entre los datos de Mortalidad por baja de Oxígeno y la concentración de oxígeno disuelto en la columna de agua

```
##  
## Pearson's product-moment correlation  
##  
## data: tidy_Mortalidad$Mort_O2 and tidy_Mortalidad$Oxigeno  
## t = -1.5747, df = 42, p-value = 0.1228  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.49807358 0.06535573  
## sample estimates:  
## cor  
## -0.2361058
```

# Hipotesis prueba de correlación

El test de Pearson, nos indica que existe una correlación negativa entre ambas variables, vale decir que a menor concentración de oxígeno en el agua aumenta la mortalidad de los peces, sin embargo el valor obtenido es muy bajo con un p-valor de 0.12 no significativo. Esto se debe a la característica de los datos utilizados para el análisis, ya que solo se consideró el promedio del ciclo productivo y no todas las mediciones durante el ciclo.

# Evaluación de supuestos

Se evalúan los siguientes supuestos:

## **Independencia**

Independencia de los datos: Las observaciones fueron obtenidas desde distintos Barrios y Centros de cultivos, por lo que existe independencia.

**Homocedasticidad** Se observa en el gráfico que las varianzas de los residuales son desiguales, por lo que no se cumple el supuesto de Homocedasticidad.

**Normalidad** la distribución de los residuales no siguen la distribución normal teórica, a pesar de que existe una porción de éstos que sí se ajusta a la línea recta diagonal teórica, por lo que podemos concluir que el supuesto de normalidad no se cumple.

# Modelo Lineal

Para la evaluación de los supuestos se utiliza el modelo lineal, donde se consideran las variables, oxígenos, barrio y N° de siembra como variables predictoras y la variable respuesta es la mortalidad por bajas de oxígeno

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	20775.0085905	1.970477e+04	1.0543134	0.2989636
Oxigeno	-2646.8731070	2.097337e+03	-1.2620163	0.2152883
Barrio16	156.7693261	1.192011e+04	0.0131517	0.9895815
Barrio17A	15994.6534490	1.081001e+04	1.4796144	0.1479216
Barrio17B	10004.8640959	1.084674e+04	0.9223845	0.3626431
Barrio7	-1366.9123705	1.149553e+04	-0.1189082	0.9060280
Barrio8	29686.5980370	1.356172e+04	2.1889995	0.0353566
Barrio9B	11342.2125493	1.299565e+04	0.8727696	0.3887372
Siembra	0.0025724	9.301900e-03	0.2765503	0.7837521

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Oxigeno	1	511562996	511562996	3.0943546	0.0873049
Barrio	6	2866232427	477705404	2.8895560	0.0214892
Siembra	1	12643790	12643790	0.0764801	0.7837521



# Conclusión

Finalmente se concluye que el modelo permite predecir como afecta la concentración de oxígeno a la variable respuesta mortalidad por oxígeno, sin embargo al revisar el  $R^2$  ajustado notamos que la capacidad predictoria es muy baja.