

Reporte de Análisis Exploratorio de Datos

Jordan I. Huanacuni

30 November 2021

Cargamos la base de datos

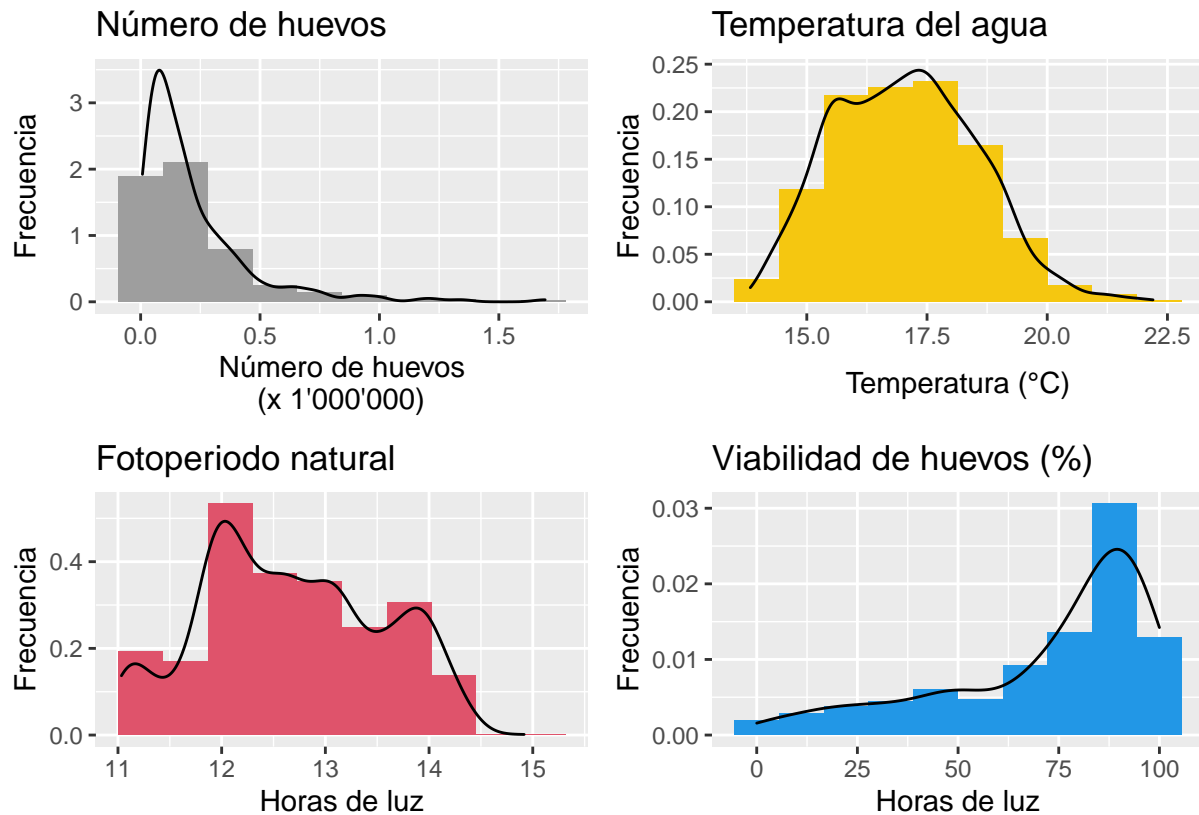
Table 1: Datos iniciales de la tabla de datos

Year	month	Temporada	Date	No_of_eggs	Egg_viability	Water_Temp	Day_length
2016	Octubre	1	2021-10-29	0.074173	80.08	19.63333	13.23333
2016	Octubre	1	2021-10-31	0.058175	68.76	18.23333	13.35000
2016	Diciembre	1	2021-12-12	0.075685	85.37	16.10000	14.13333
2016	Diciembre	1	2021-12-17	0.309222	99.00	18.73333	14.08333
2016	Diciembre	1	2021-12-18	0.124297	86.64	17.85000	14.23333
2016	Diciembre	1	2021-12-20	0.729363	95.50	17.86667	14.28333

Table 2: Datos finales de la tabla de datos

Year	month	Temporada	Date	No_of_eggs	Egg_viability	Water_Temp	Day_length
2020	Diciembre	5	2021-12-11	0.419130	90.4	16.93333	13.18333
2020	Diciembre	5	2021-12-13	0.219478	74.4	16.86667	13.20000
2020	Diciembre	5	2021-12-19	0.327930	91.2	18.40000	13.20000
2020	Diciembre	5	2021-12-20	0.474782	91.5	18.80000	13.21667
2020	Diciembre	5	2021-12-21	0.300069	86.9	18.80000	13.20000
2020	Diciembre	5	2021-12-25	0.537391	77.6	17.73333	13.20000

Variación de variables con histogramas



COMENTARIO El histograma de los datos de número de huevos indica una distribución asimétrica positiva.

El histograma de los datos de temperatura del agua, aparentemente tienen una distribución normal.

El histograma de los datos de las horas luz del ambiente, no tienen una distribución muy confusa, aparentemente es normal.

El histograma de los datos de viabilidad de huevos tienen una distribución asimétrica negativa.

Identificación de datos balanceados

Eventos de desove por temporada

Table 3: Cantidad de desoves por temporada

Temporada	Número de desoves
1	93
2	54
3	56
4	57
5	25

COMENTARIO: La cantidad de desoves clasificados según temporada, no están balanceados

Días evaluados por Año

Table 4: Cantidad de días con registro de temperatura y fotoperiodo

Año	Días evaluados
2016	366
2017	365
2018	365
2019	365
2020	366
2021	365

COMENTARIO: La cantidad de datos de temperatura y fotoperiodo natural registrados durante los años difieren en 1 observación, debido a los años bisiestos, no están balanceados.

Días evaluados por mes

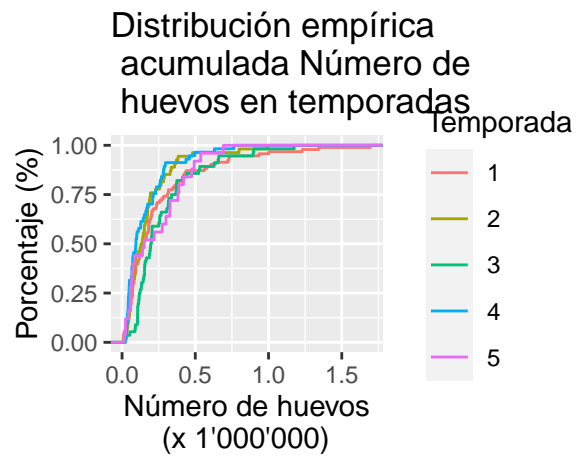
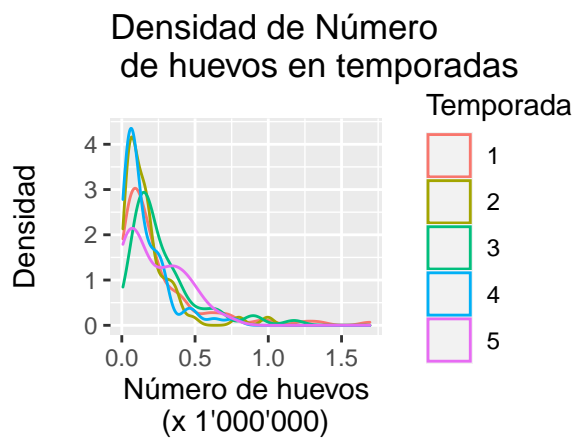
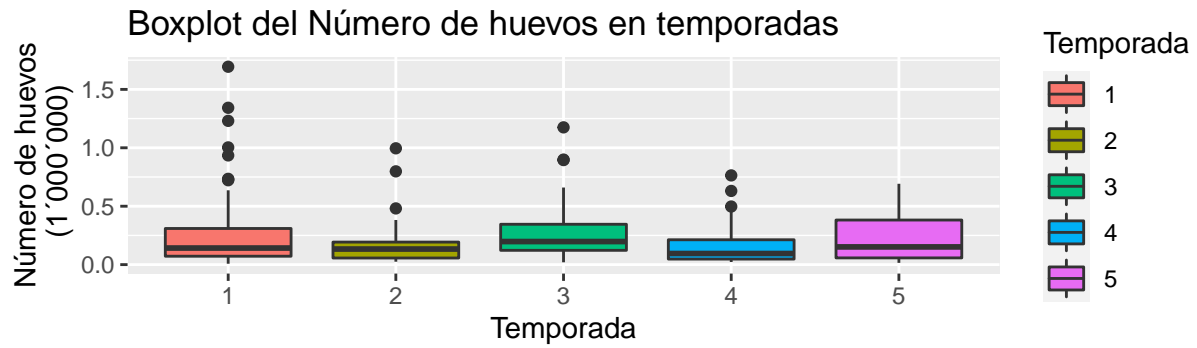
Table 5: Cantidad de días por mes con registro de temperatura y fotoperiodo

Mes	Días evaluados
Enero	186
Febrero	170
Marzo	186
Abril	180
Mayo	186
Junio	180
Julio	186
Agosto	186
Setiembre	180
Octubre	186
Noviembre	180
Diciembre	186

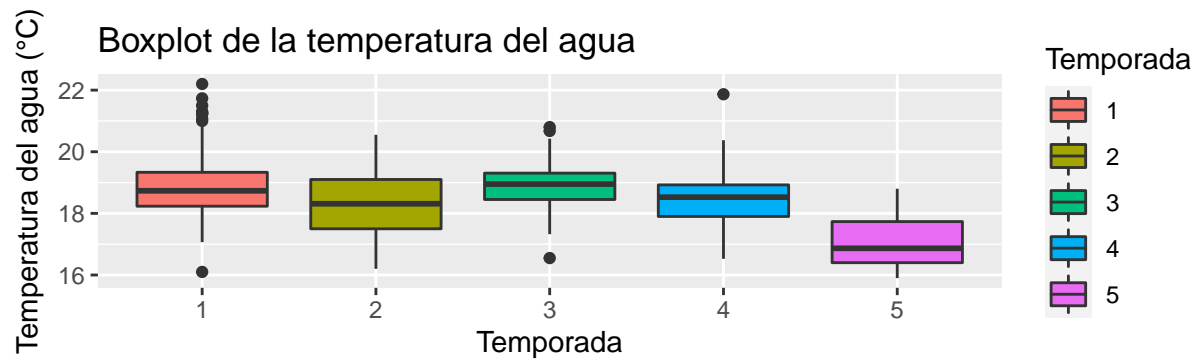
COMENTARIO: La cantidad de datos de temperatura y fotoperiodo natural registrados durante los meses difieren en 1 observación, debido a los años bisiestos, no están balanceados.

Gráficos exploratorios de Relación entre variables

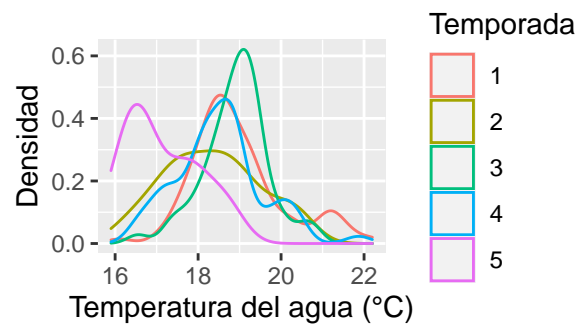
Desoves por temporada



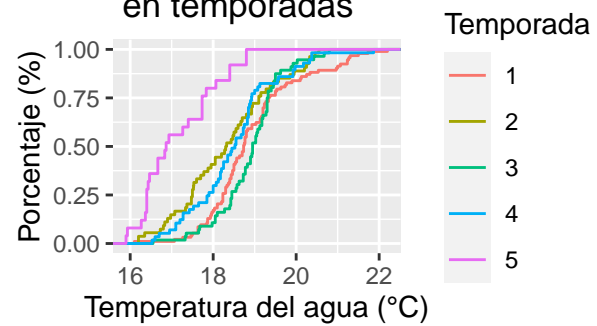
Temperatura del agua por temporada



Temperatura del agua en temporadas

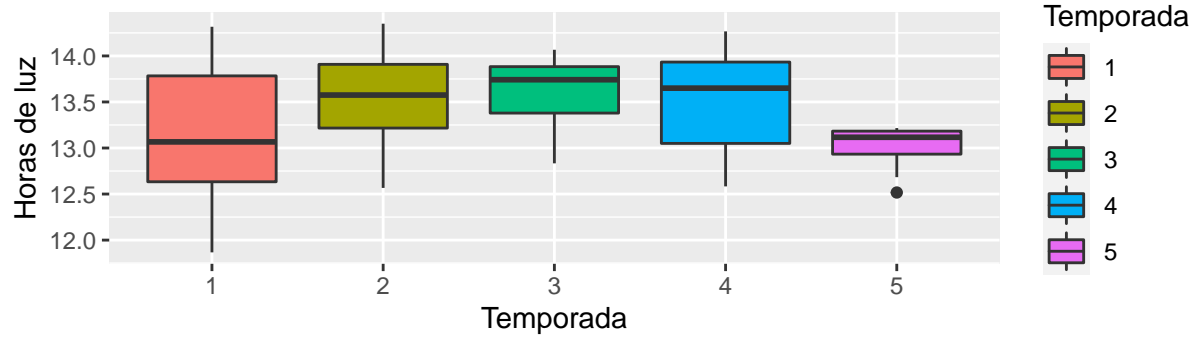


Distribución empírica acumulada de las temperaturas del agua en temporadas

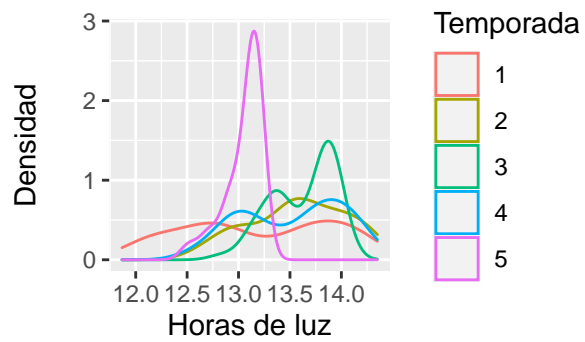


Horas de luz por temporada

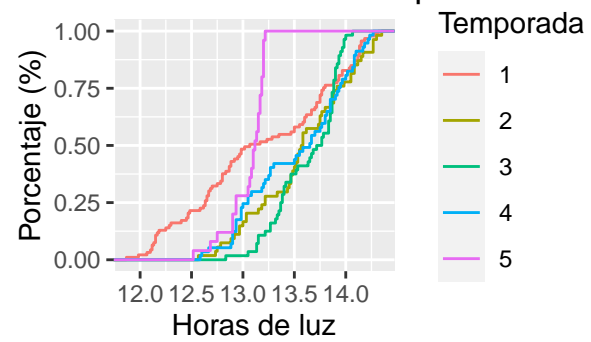
Boxplot del Fotoperiodo natural en temporadas



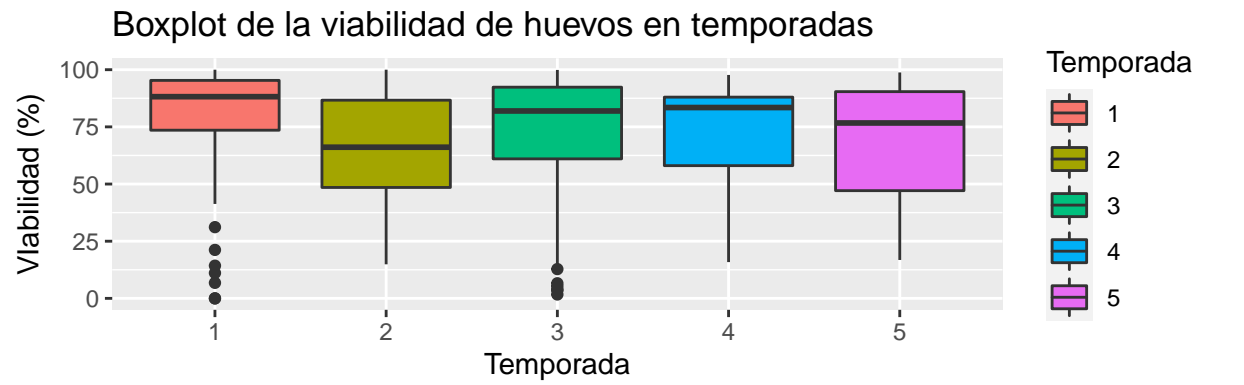
Densidad del fotoperiodo natural



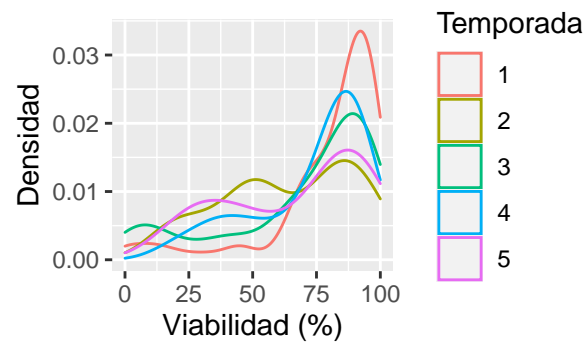
Distribución empírica acumulada del fotoperiodo natural



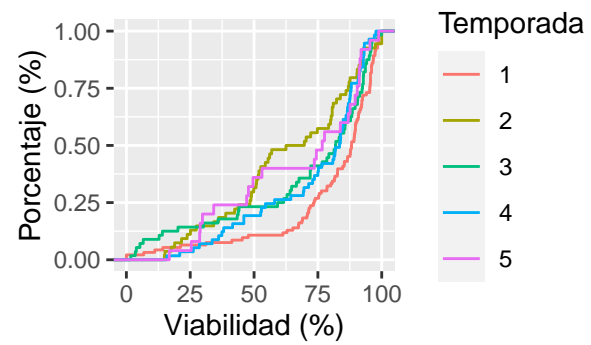
Viabilidad de huevos por temporada



Densidad de la viabilidad de hue

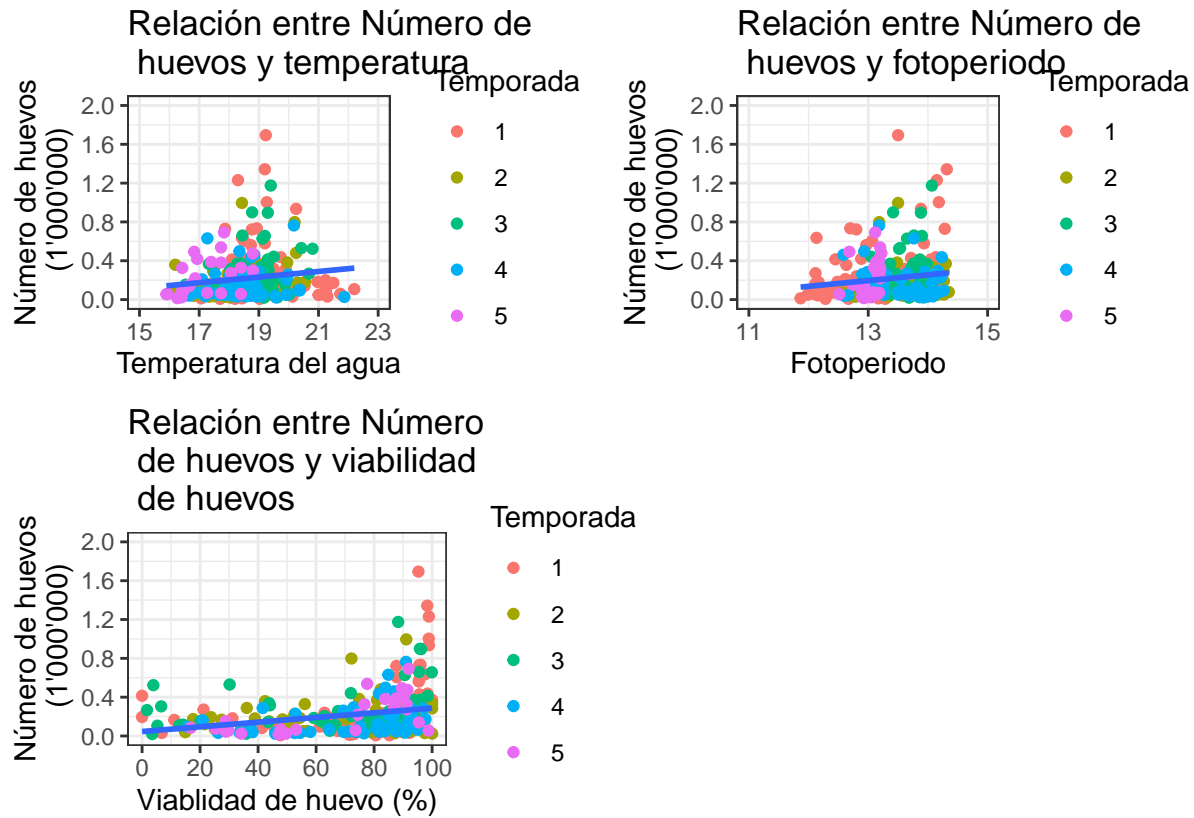


Distribución empírica acumulada de la viabilidad de hu



Relación: Número de huevos vs Temperatura/Fotoperiodo/Viabilidad

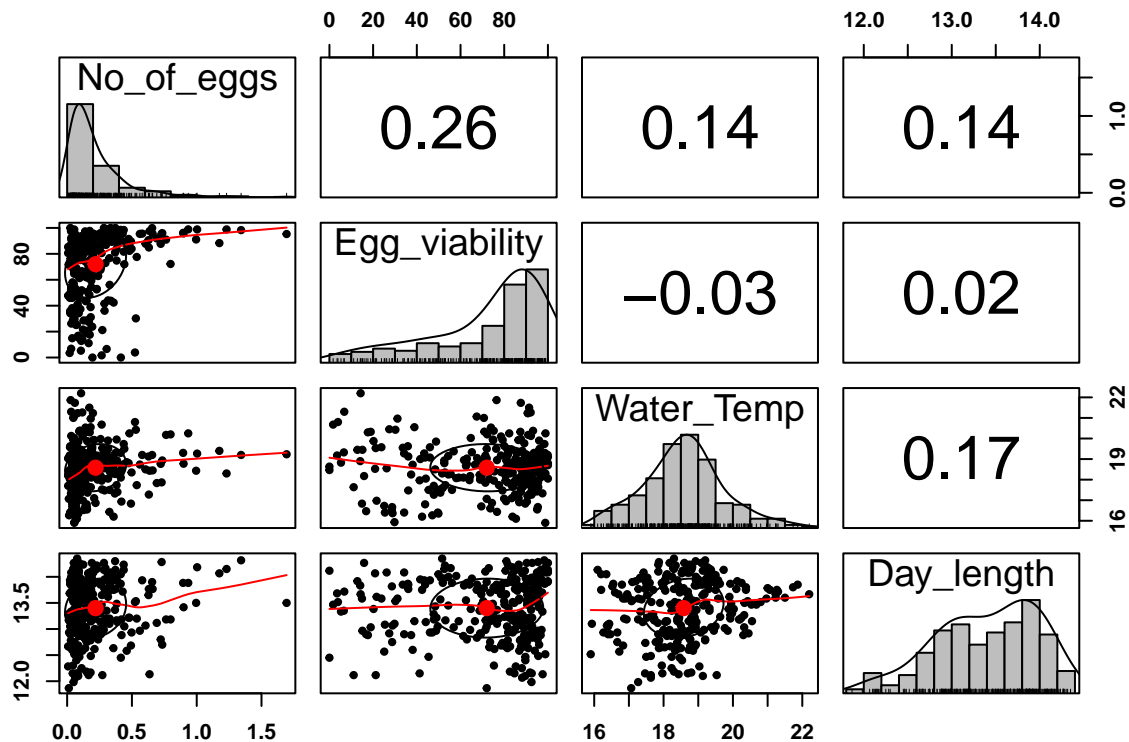
Gráfico de puntos del Número de huevos vs Temperatura del agua



COMENTARIO: Según los gráficos, los datos se encuentran muy dispersos, sin una aparente relación entre las variables (Temperatura, Fotoperiodo y Viabilidad) con la cantidad de huevos de sargo.

Gráfica de correlación de variables continuas

Número de huevos vs Temperatura/Fotoperiodo/Viabilidad



COMENTARIO: Según el gráfico, las variables tienen una correlación muy baja. Número de huevos vs viabilidad: 0.26

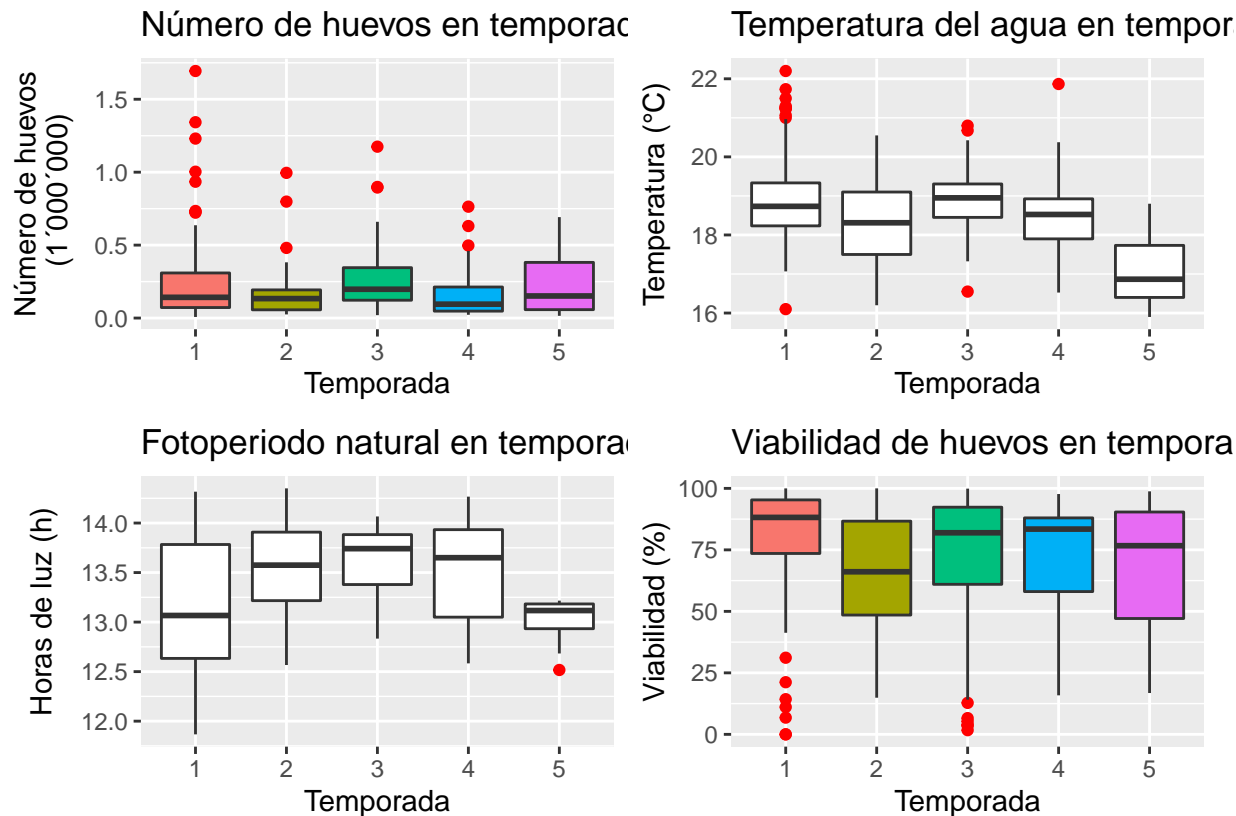
Número de huevos vs Temperatura del agua: 0.14

Número de huevos vs Fotoperiodo natural: 0.14

lo cual indica que no hay una relación correlación positiva débil entre las variables evaluadas.

Identificación de datos atípicos

Número de huevos



COMENTARIO El número de huevos presenta datos atípicos en las temporadas 1, 2, 3 y 4.

La temperatura del agua presenta datos atípicos en las temporadas 1, 3 y 4.

El fotoperiodo natural presenta 1 dato atípico en las temporadas 5.

La viabilidad de desoves presenta datos atípicos en las temporadas 1 y 2.

Resumen de datos y estadística descriptiva usando Tablas

Resumen estadístico de huevos por temporada

Table 6: Resumen de cantidad de huevos (x 1'000'000) agrupado por temporada

Temporada	mean	sd	var	max	min	n
1	0.2486032	0.3001851	0.0901111	1.693808	0.007569	93
2	0.1754970	0.1796802	0.0322850	0.995483	0.024923	54
3	0.2756932	0.2310949	0.0534049	1.174991	0.019695	56

Temporada	mean	sd	var	max	min	n
4	0.1549502	0.1545281	0.0238789	0.763652	0.022260	57
5	0.2278548	0.1951835	0.0380966	0.692000	0.015130	25
NA	NA	NA	NA	NA	NA	1907

Resumen de Temperaturas por temporada

Table 7: Resumen de la temperatura del agua (°C) agrupado por temporadas

Temporada	mean	sd	var	max	min	n
1	18.94615	1.1208246	1.2562478	22.20000	16.100	93
2	18.33873	1.1591211	1.3435617	20.55000	16.200	54
3	18.90997	0.7898351	0.6238394	20.80000	16.550	56
4	18.51491	1.0564601	1.1161080	21.86667	16.525	57
5	17.14800	0.8844375	0.7822296	18.80000	15.900	25
NA	NA	NA	NA	NA	NA	1907

Resumen de Fotoperiodo natural por temporada

Table 8: Resumen de las horas de luz del fotoperiodo natural agrupado por meses

Temporada	mean	sd	var	max	min	n
1	13.18423	0.7141951	0.5100747	14.31667	11.86667	93
2	13.57315	0.4689457	0.2199100	14.35000	12.56667	54
3	13.62827	0.3061808	0.0937467	14.06667	12.83333	56
4	13.51813	0.4906495	0.2407369	14.26667	12.58333	57
5	13.05133	0.1825057	0.0333083	13.21667	12.51667	25
NA	NA	NA	NA	NA	NA	1907

SEGUNDA PARTE DE LA PRUEBA

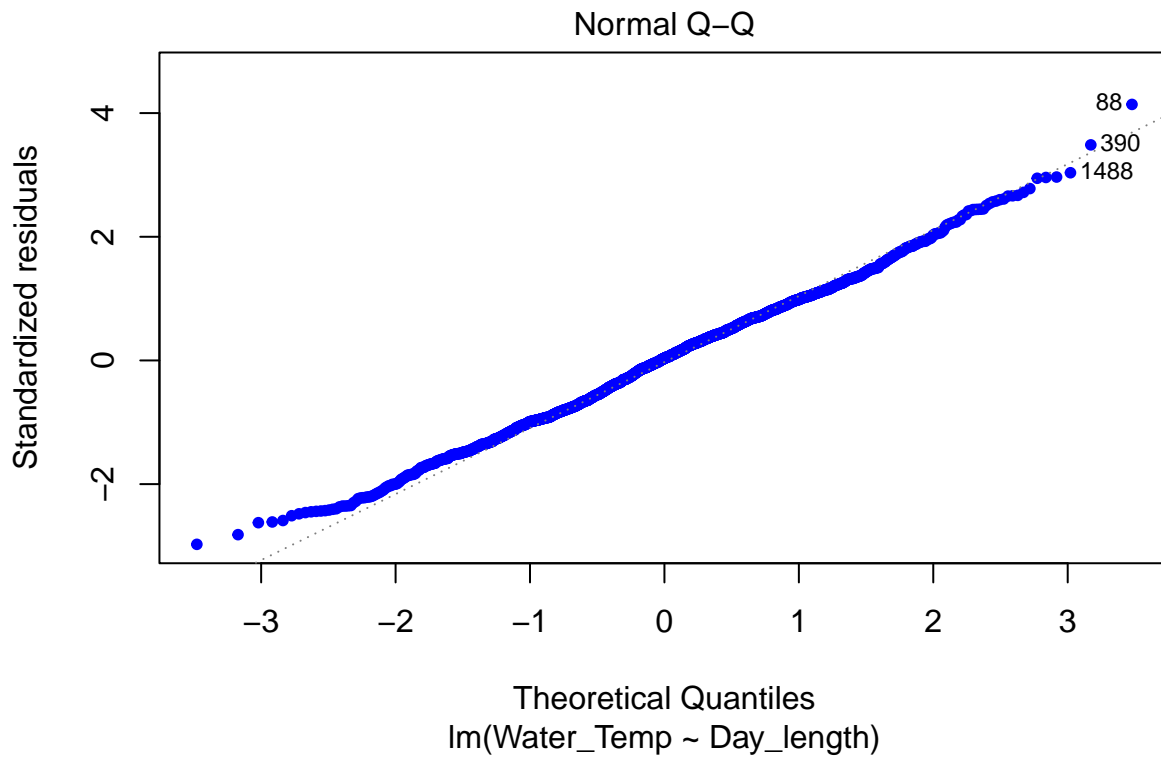
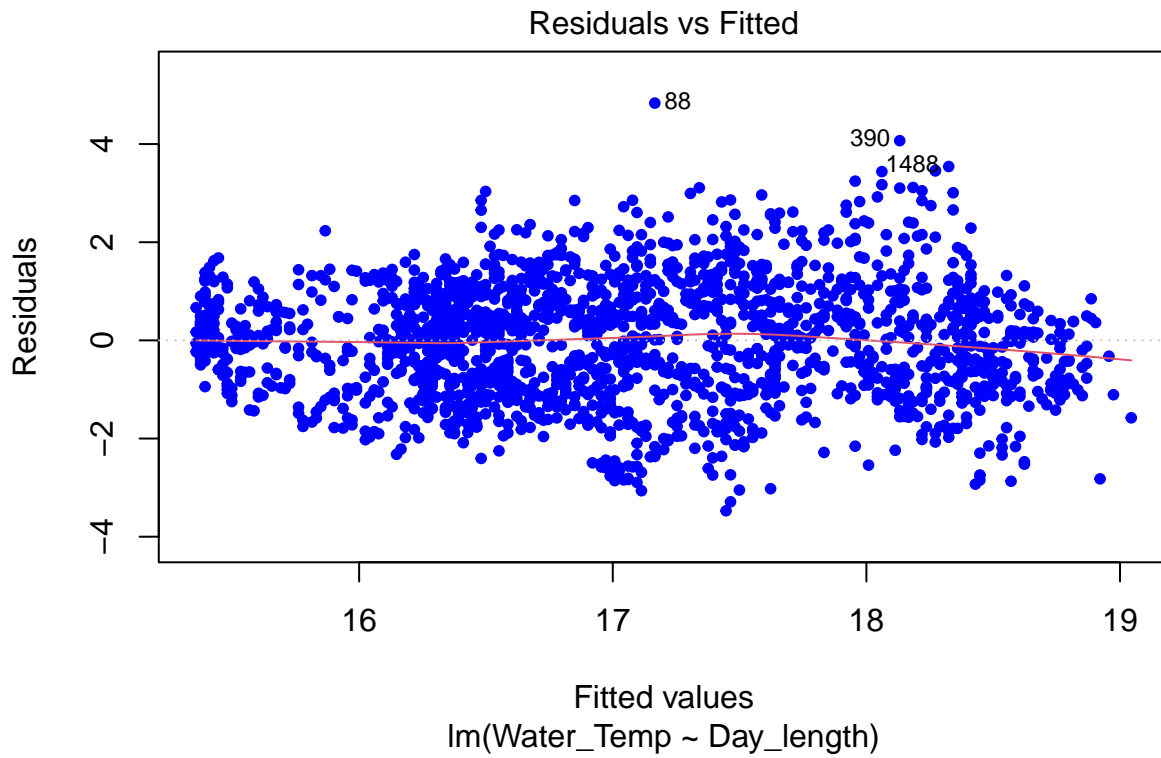
Objetivo

Determinar el modelo estadístico que se ajusta a la relación de temperatura del agua con crecimiento de larvas de *Anisotremus scapularis* *Paralichthys adspersus* “lenguado fino”

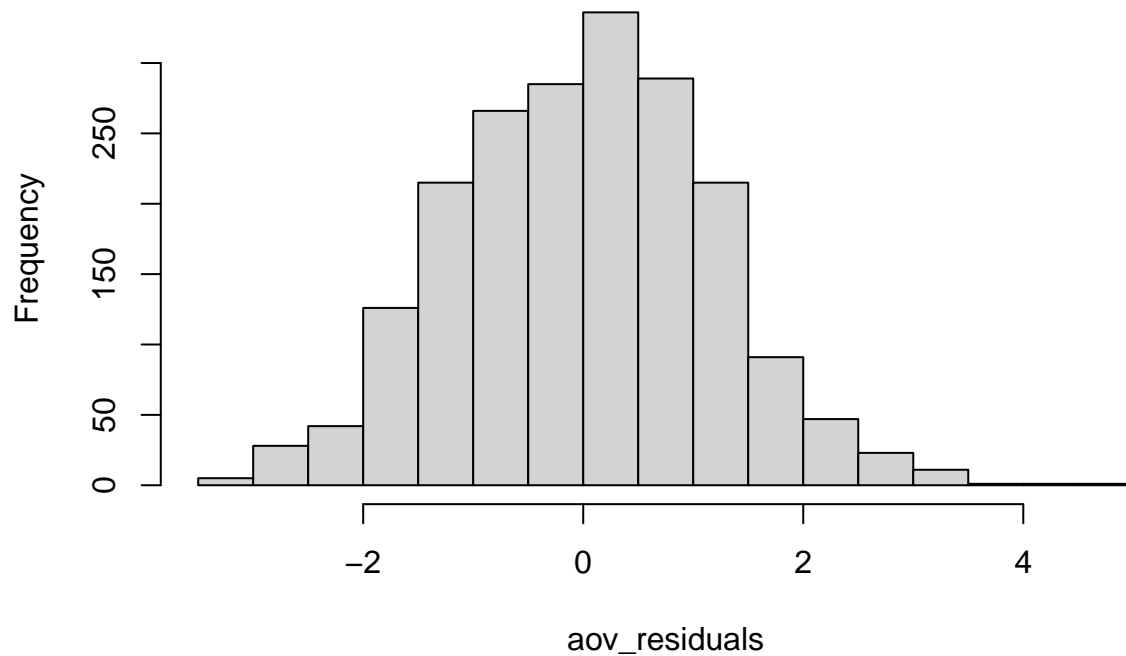
Supuesto de Normalidad

```
##
## Call:
## lm(formula = Water_Temp ~ Day_length, data = Sargo)
```

```
##
## Coefficients:
## (Intercept)  Day_length
##      3.735      1.053
```



Histograma de residuales



```
##  
## Shapiro-Wilk normality test  
##  
## data: aov_residuales  
## W = 0.99764, p-value = 0.004738
```

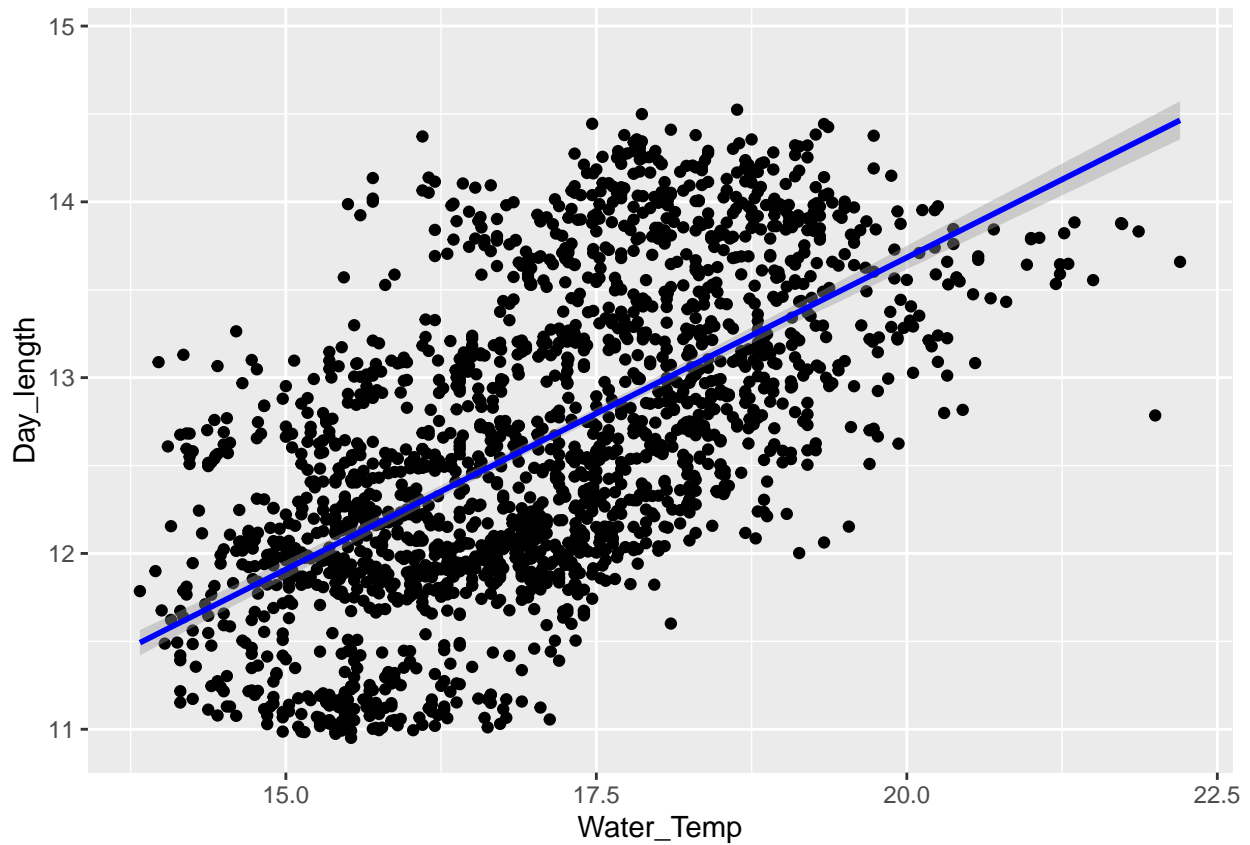
La prueba de Shapiro nos indica que no hay normalidad entre los datos (p-value = 0.04738)

Comparacion de regresion lineal y no lineal

Modelo 1. Regresion lineal simple

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.734712	0.3880315	9.624766	0
Day_length	1.053421	0.0306412	34.379236	0

```
## Warning: Removed 210 rows containing non-finite values (stat_smooth).  
## Warning: Removed 210 rows containing missing values (geom_point).
```

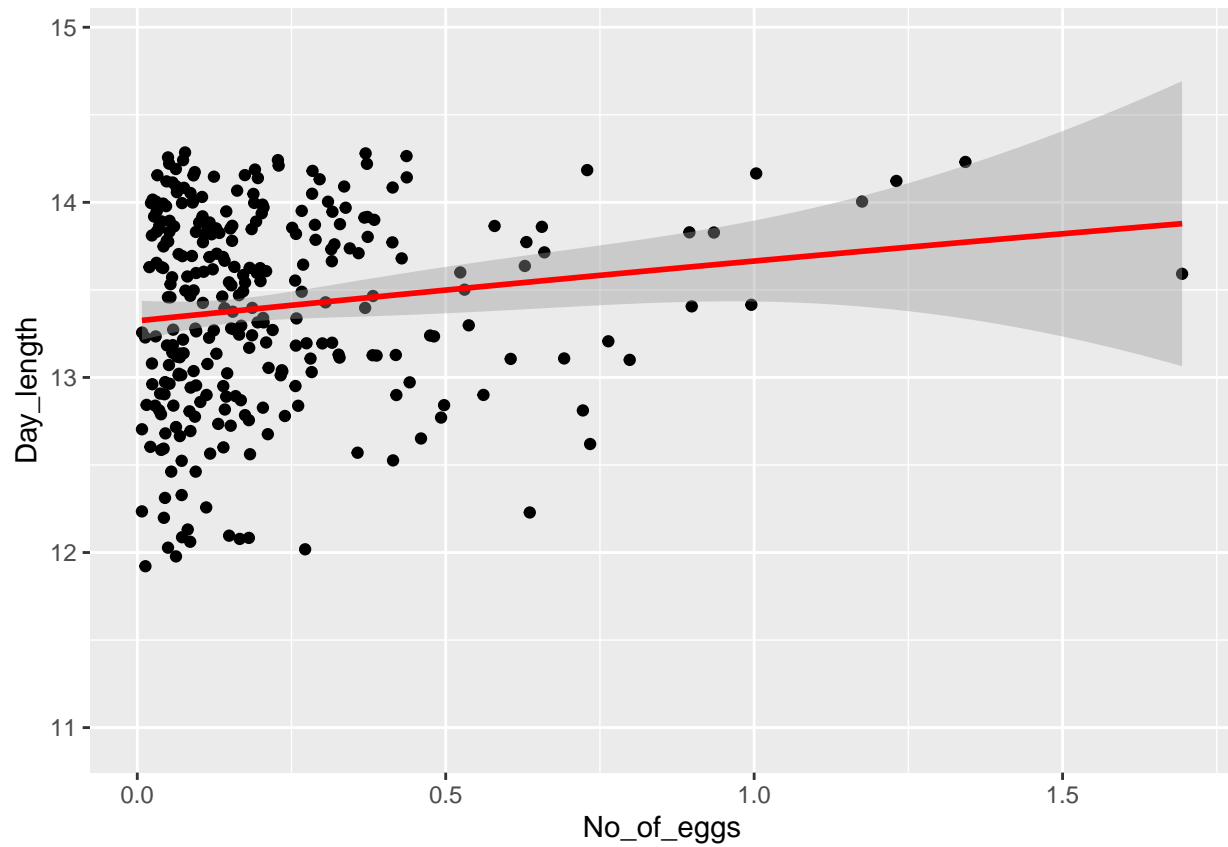


Descripcion: Segun la figura los datos no interceptan con la pendiente, por lo tanto, no presentan un modelo lineal

Modelo 2. Regresion No lineal

```
## Warning: Removed 1907 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1907 rows containing missing values (geom_point).
```



Los datos no se ajustan a un modelo no lineal

Conclusión

Los datos de Temperatura y fotoperiodo natural, no se adecuan a un modelo lineal, es necesario la inclusion de mas variables para una adecuada evaluacion.