Tecnológico de Monterrey

Actividad:

Reporte Final: Los peces y el mercurio

Módulo:

Módulo 1: Estadística e Inteligencia artificial avanzada para la ciencia de datos

Grupo:

TC3006C.101

Nombre:

Franco Quintanilla Fuentes - A00826953

Maestra:

Blanca R. Ruiz Hernández

Fecha:

14 de septiembre de 2022

Resumen

La contaminación por mercurio de peces en el agua dulce comestibles es una amenaza directa contra nuestra salud, por lo que realizamos un estudio estadístico con el fin de determinar las variables que influyen a la contaminación por mercurio. Para poder responder las preguntas, utilizamos métodos estadísticos como: Regresión lineal simple, y Análisis de la Varianza (ANOVA). En donde llegamos a los resultados que el PH es un factor que ayuda a reducir la contaminación de mercurio en los peces, y que la concentración de mercurio no varía entre las edades de los peces.

Introducción

En la investigación de (OCU, 2021) se menciona que el mercurio se libera al medio ambiente a través de procesos naturales, el cual se presenta en el suelo, el agua y la atmósfera. El problema empieza cuando el humano aporta grandes cantidades de mercurio al medio ambiente a través de distintos procesos. Como repercusión, las grandes concentraciones de mercurio llegan al agua, por ende a los animales, que en nuestro caso a considerar, los peces.

A nosotros nos interesa hacer un estudio estadístico, ya que el mercurio puede ocasionar efectos tóxicos en tanto órganos como sistemas, algunos ejemplos son:

- El sistema nervioso
- Los riñones
- El hígado
- Los órganos reproductivos

Los daños más peligrosos son los neurotóxicos, ya que sus efectos repercuten sobre el desarrollo neuronal, y los períodos de exposición durante el embarazo. Por eso, nuestro estudio se basó en una pregunta base, la cual fue desglosando 2 preguntas consecuentes.

- ¿Cuáles son los principales factores que influyen en el nivel de contaminación por mercurio en los peces de los lagos de Florida?

Las preguntas subyacentes de la pregunta base, son las que más énfasis tienen a la hora de preocuparnos por la salud humana.

1. ¿Las concentraciones de alcalinidad, clorofila, calcio en el agua del lago influyen en la concentración de mercurio de los peces?
2. ¿Habrá diferencia significativa entre la concentración de mercurio por la edad de los peces?

Con estas preguntas podemos empezar a realizar nuestro análisis estadístico, para poder sacar resultados educados y sustentados con estadística.

Análisis de resultados

Regresión Lineal Simple

Para poder obtener resultados, primero tenemos que hacer un análisis de los datos, ver su distribución, su comportamiento, sus datos atípicos, su media, etc.
En nuestro caso, una de las primeras cosas que hicimos, fue limpiar los datos, para después poder hacer una matriz de correlación para observar cuál factor influye más con nuestra variable objetivo, que en este caso era el ***Mínimo de la concentración de mercurio*** (X9).

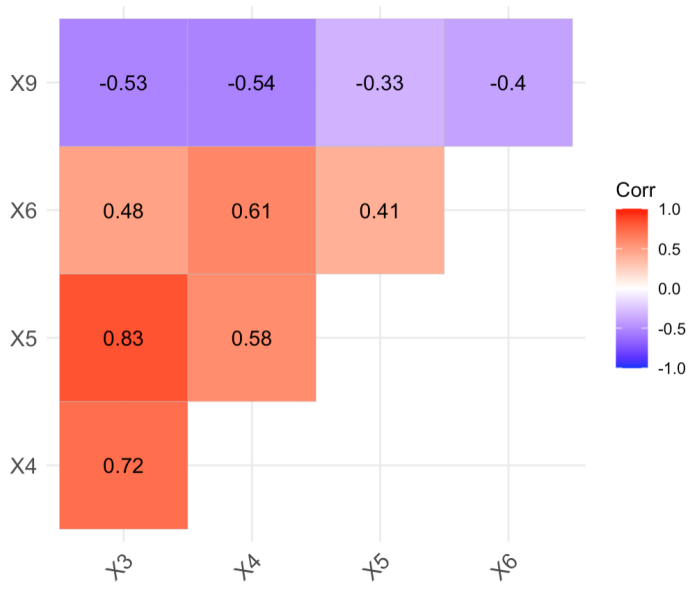

Figura 1: Matriz de correlación de los principales factores.

En la figura 1, podemos observar que X4, que en este caso es el PH es el factor con mayor correlatividad en base a nuestra variable objetivo, con un valor de -0.54, por lo cual escogimos el ***PH*** (X4) como variable independiente para nuestro estudio y modelaje de la regresión lineal simple.

Como ya contamos con una variable objetivo, podemos visualizar sus datos, en donde usamos un boxplot para ver sus cuartiles y sus datos atípicos, así como un histograma para ver la distribución de los mismos.

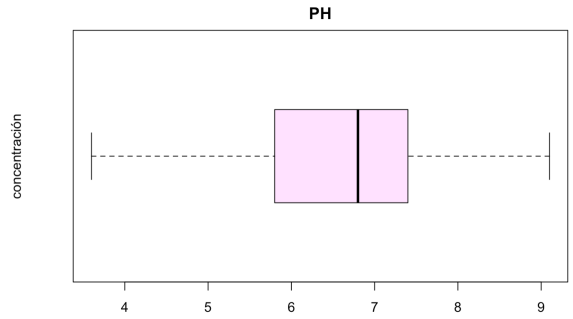

Figura 2a: Box plot del PH

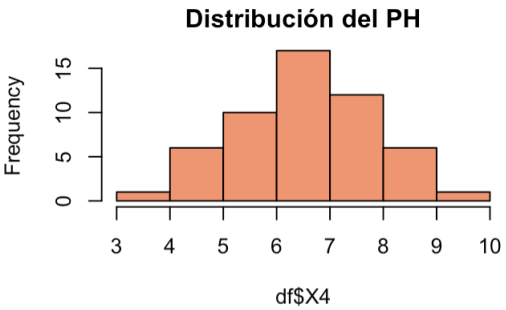

Figura 2b: Distribución del PH

Figura 2: Visualización de la concentración del PH y su distribución.

Como podemos observar en la figura 2a, los datos de la concentración del PH vienen “limpios” ya que no cuenta con ningún dato atípico y todos sus datos se encuentran dentro de su primer y cuarto cuartil. En la figura 2b podemos observar que los datos tienen un comportamiento Normal, ya que el mismo histograma se comporta de esa manera, lo que también nos ayudaría si quisiéramos hacer una prueba de hipótesis con la misma variable. En nuestro caso nos ayuda a visualizar nuestro modelo con el análisis de los residuos.

Hecho esto, podemos observar el comportamiento entre ambas de nuestras variables, es decir, entre el Mínimo de la concentración de mercurio, y la concentración de PH. Por lo que optamos por hacer un scatter plot de los mismos, para tener una mejor referencia de la posible tendencia que los datos pudieran seguir.

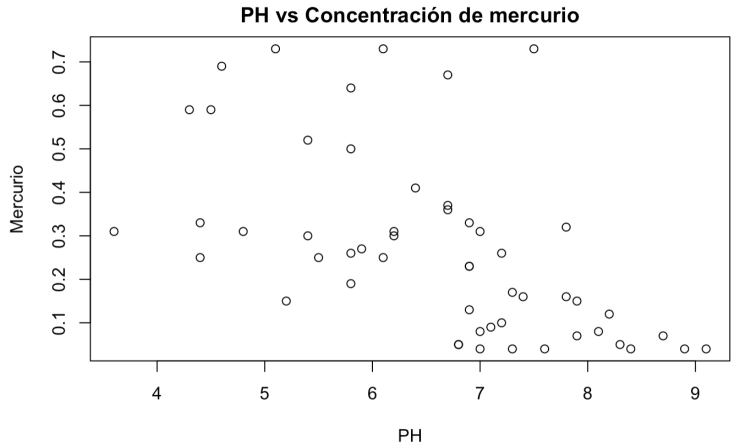

Figura 3: Comportamiento del PH vs la Concentración de mercurio.

Como podemos observar, los datos siguen una tendencia linealmente negativa, en donde nos dicen que entre mayor sea el PH, la concentración de mercurio va a disminuir.

Con esto, el siguiente paso es hacer nuestra regresión lineal simple, en donde con ayuda de R, nuestra ecuación nos quedó de la siguiente manera.

$$E(Y) = 0.878 - 0.0919\,x,$$

La cual nos explica que el valor esperado de la concentración de mercurio, cuando el valor del PH es 0, la concentración de mercurio inicial va a ser de 0.878, y que entre mayor sea la concentración de PH en el agua, el valor esperado de la concentración de mercurio va a disminuir en un factor de 0.091. Lo que nos dice lo que explicamos anteriormente, que entre mayor sea el PH, la concentración de mercurio va a disminuir.

Por otra parte, también sacamos los intervalos de confianza para delimitar con mayor presión el valor esperado, con un intervalo de confianza del 95%, lo cual nos dio los siguientes resultados.

```
                2.5 %      97.5 %
(Intercept)  0.6259342  1.13111675
x           -0.1295306 -0.05427618
```

Figura 4: Intervalos de confianza con función “*confint*” de R.

Estos datos así curados son muy difíciles de visualizar, por lo que ahora vamos a mostrar nuestro modelo de regresión lineal simple, con los intervalos de confianza, nos da la siguiente gráfica.

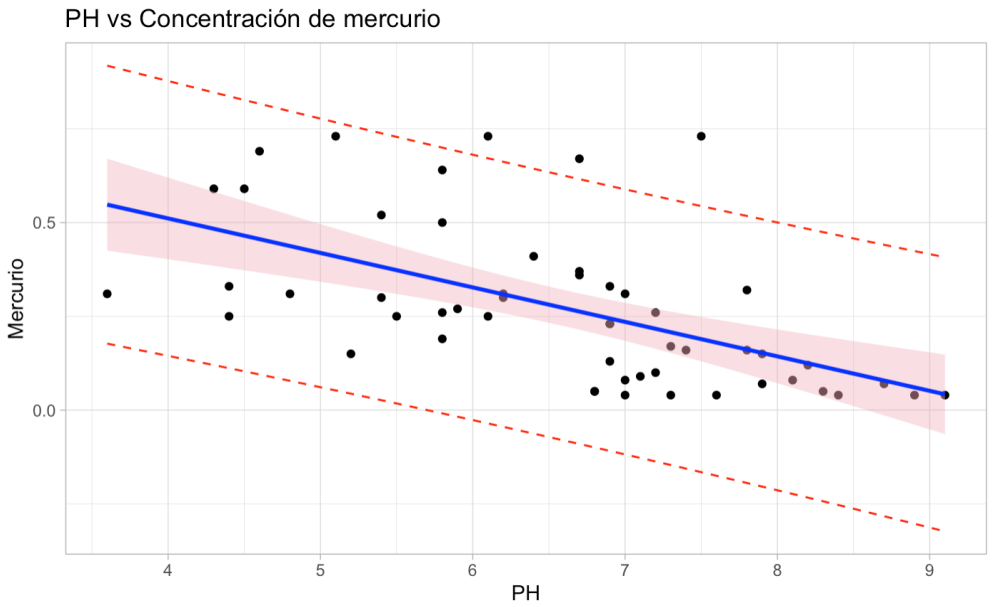

Figura 5: Modelo de regresión lineal simple.

Como podemos observar nuestro modelo se comporta de buena manera, aunque no es tan exacta, esto quiere decir que nuestro modelo se ajusta a la tendencia la cual siguen los datos, pero hay mucha variabilidad en los mismos, por lo cual es muy difícil poder hacer una predicción acertada. También afecta a que los intervalos de confianza de nuestro modelo son muy pequeños, esto no le da al modelo tanto rango de valores aceptables.

ANOVA (Analysis of variance)

Para el caso del ANOVA, como se deben usar variables categóricas, cambiamos nuestra variable independiente por ***La edad de los peces*** (X12), en donde con ayuda de R que hace todo el análisis, obtuvimos el siguiente resultado.

```
            Df Sum Sq Mean Sq F value Pr(>F)
edad         1 0.0427 0.04266   0.974  0.328
Residuals   51 2.2330 0.04378
```

Figura 6: ANOVA en R.

Este análisis del ANOVA nos dice mucho, pero a su vez tenemos que hacer todo un Análisis para cada tipo de edad de los peces (***0: Jóvenes, 1: Maduros***), en donde obtuvimos los siguientes resultados importantes.

	Jóvenes	***Maduros***
Media de la concentración de mercurio por la edad de los peces	0.214	0.286
Desviación estándar de la concentración de mercurio por la edad de los peces	0.214	0.208
Tamaño de la muestra de la concentración de mercurio por la edad de los peces	10	43

Tabla 1: Análisis estadístico para cada tipo de edad de los peces.

Con este análisis, podemos obtener los intervalos de confianza para cada tipo de edad de los peces, en donde lo más óptimo es graficarlos para tener una mejor representación.

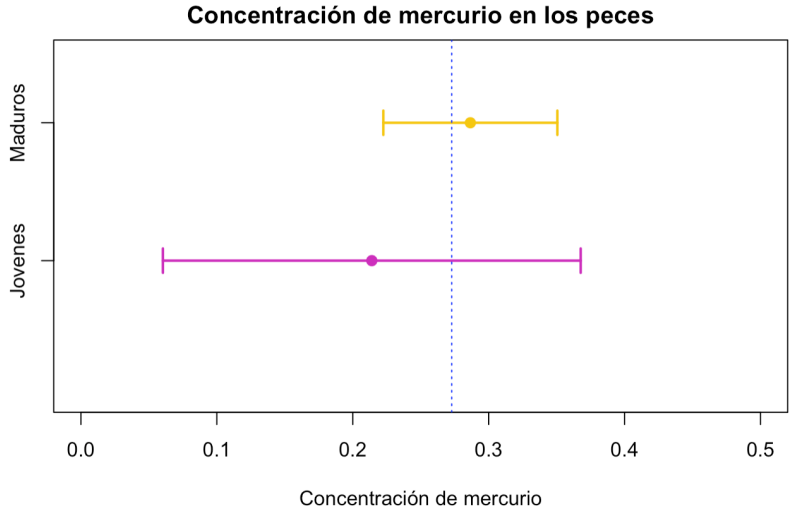

Figura 7: Intervalos de confianza de la concentración de mercurio por cada tipo de edad de los peces.

Desde aquí podemos observar que la concentración de mercurio por la edad de los peces no afecta, ya que una contiene a la otra en los intervalos de confianza, por lo que se podría decir que son iguales, donde además, ambas se encuentran dentro de la media poblacional.

Conclusión

Después de hacer todo el análisis estadístico en base a la concentración de mercurio en los peces, podemos observar que:

- Gracias a la regresión lineal simple, podemos ver que cuando un lago tiene mayor PH, es decir, el agua es más alcalina, la concentración de mercurio en los peces va a disminuir, por el contrario si el agua es más ácida, entonces la concentración de mercurio en los peces va a ser mayor.

- Después de hacer el análisis de varianza en base a la edad de los peces y su concentración de mercurio, podemos responder otra de las preguntas, y es que NO habrá diferencia significativa entre la concentración de mercurio por la edad de los peces, y esto lo podemos demostrar gracias a los intervalos de confianza.

Referencias

OCU. (2021, 6 abril). Mercurio en el pescado. www.ocu.org. Recuperado 14 de septiembre de 2022, de https://www.ocu.org/alimentacion/alimentos/noticias/mercurio-en-pescado-un-problema-serio522454

Anexos

Repositorio de Github

https://github.com/francoquintanilla0/Mercurio-Peces

Regresión Lineal Simple

Análisis de los residuos

Normalidad de los residuos

Para la prueba de normalidad de residuos, hacemos una prueba de hipótesis con el shapiro test, pero como es difícil de interpretarlo, agregaremos sus respectivas gráficas.

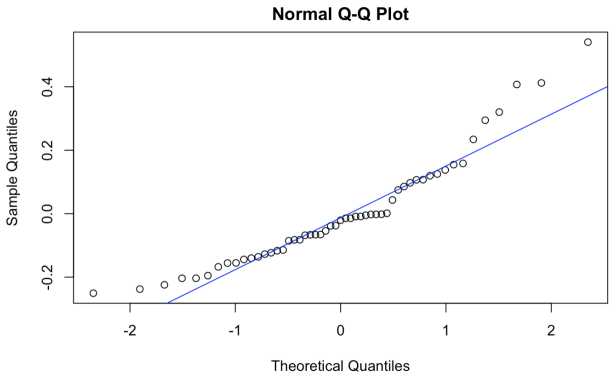

Figura 8: Prueba de normalidad: QQ-Plot.

Como podemos ver, el modelo se comporta como una distribución con colas gruesas es decir, que tiene baja curtosis, del tipo platicúrtica, pero que también tiene una asimetría positiva es decir que su sesgo va a la derecha.

Otra manera de visualizar los datos, es con la ayuda de un histograma, lo cual es lo siguiente que vamos a hacer.

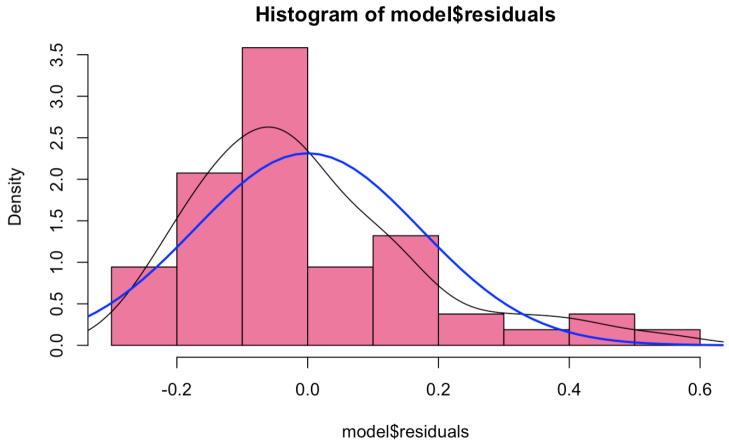

Figura 9: Prueba de normalidad: Histograma.

Con el histograma podemos ver el comportamiento que concluimos con la ayuda de las pruebas de normalidad, que nuestro modelo cuenta con una curtosis platicúrtica, y lo más visual es que tiene una asimetría positiva lo que nos dice que su sesgo va a la derecha.

Verificación de media cero

Para la verificación de la media cero, vamos a utilizar el t.test de R, en donde nos dio que el valor de $t^* = 3.6434e-16$ a comparación de t_0, es que t^* es muy cercano a 0, por lo que está dentro de los límites para poder aceptar la hipótesis de que los errores se comportan como una normal.

Homocedasticidad

Con la homocedasticidad, buscamos que los residuos de nuestro modelo no sigan ninguna tendencia, ni que cuenten con heterocedasticidad, por lo que tenemos que graficar los residuos

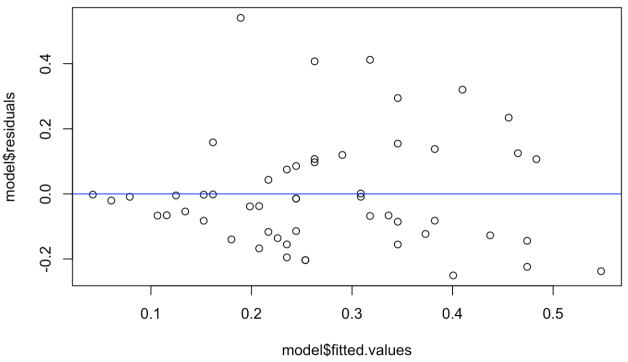

Figura 10: Prueba de homocedasticidad con los residuos.

Con este análisis de la homocedasticidad nos damos cuenta de una buena vez por todas que nuestro modelo es el adecuado para definir la comparación entre el PH y la concentración de mercurio, ya que se ve que el modelo cuenta con homocedasticidad, y sus residuos no cuentan con una tendencia específica.

ANOVA

Verificación de supuestos

Independencia

Para el caso de la ANOVA, nos interesa más ver la independencia de nuestro modelo, por lo que graficamos los residuos con el orden de la observación, y de la misma manera, buscamos que esta no siga una tendencia en específico, sino un comportamiento residual.

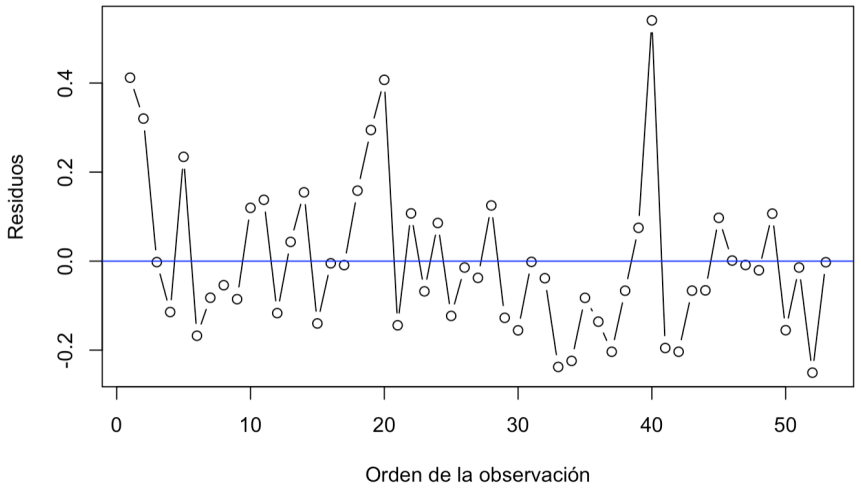

Figura 11: Prueba de Independencia.

Después de hacer el análisis de las varianzas, podemos observar que la edad de los peces, ya sean jóvenes o maduros, no afecta en la concentración de mercurio en los peces de los distintos lagos. Aunque pensemos que esto pueda estar influenciado ya que la cantidad de peces en la muestra favorece a los peces maduros, ya que son mucho más la cantidad que los peces jóvenes, pero aun con esta diferencia podemos observar que las concentraciones de mercurio son las mismas.