

Reporte Mercurio



Grupo: 102

Módulo 1: Estadística para ciencia de datos e Inteligencia artificial avanzada para la ciencia de datos I (Gpo 102)

Iker Ledesma Durán

A01653115

18/09/2021.

Reporte Mercurio.

Resumen.

En una base de datos se registraron en total 53 lagos, de estos se tomaron en cuenta 12 variables. Estas 12 variables toman en cuenta distintos factores, incluyendo poblaciones de peces estudiadas, para saber si las concentraciones de mercurio son alarmantes e incluso dañinas para la vida humana.

Introducción.

Para este trabajo, se utilizarán dos métodos distintos, el primero será una regresión lineal múltiple y el segundo será una prueba de hipótesis, estos métodos serán usados con la finalidad de poder saber si las métricas registradas en la base de datos presentan valores que muestren una amenaza para la vida digna y la potabilidad del agua. Primeramente, se realizó un análisis exploratorio para que se conociera más que nada los valores atípicos que hay en los datos, estudiar bien qué es lo que nos presentan las variables y finalmente saber la correlación que existe entre las variables. Posteriormente se aplicarán los métodos anteriormente mencionados para poder llegar a una conclusión.

Análisis de Resultados.

Para abrir está la matriz de correlación:

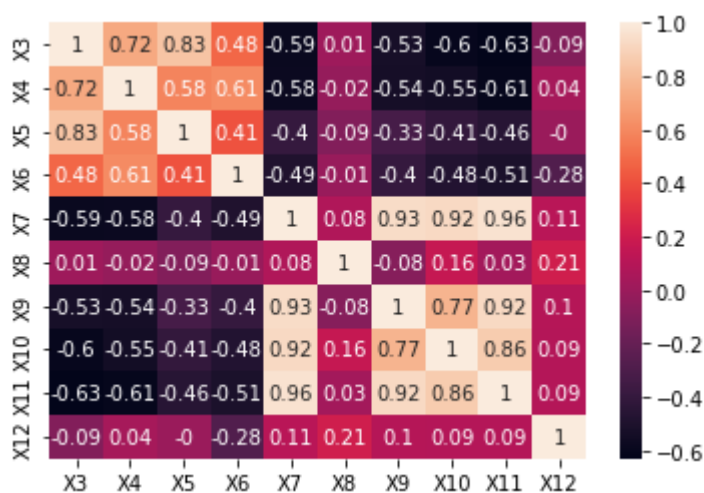


Figura 1: Matriz de correlación.

En esta podemos ver una fuerte correlación en la variable X7 con las variables X9, X10 y X11, por lo que muestra ser bastante útil para la utilización de modelos.

Para el primer modelo, regresión lineal, se tomó como variable objetivo la variable X7 ya que era la que más altas correlaciones presentaba con respecto a variables relacionadas a la población de pescados, por lo que si presentaba un riesgo para los peces que habitaban esos lagos por ende también iba a presentar un riesgo para la vida humana que dependa de ellos.

Como primer estimador se encontró que la intercepta es igual a un valor de -0.0746, por lo que se podría intuir que hay una concentración media de negativa de mercurio.

Posteriormente, se encontró que el valor del coeficiente de determinación es de 0.9742, por lo que se puede asumir que hay una alta precisión en el modelo.

Por último, se mostrarán los pasos seguidos para la realización de la prueba de hipótesis:

1. Hipótesis

- $H_0 : P = 0.5$
- $H_1 : P < 0.5$

2. Regla de decisión

```
z0 = norm.ppf(.01)
z0
-2.3263478740408408
```

Rechazamos H_0 si:

- Si $|z^*| > 2.326$
- Si valor $p < 0.01$

Figura 2: Hipótesis y Regla de decisión.

3. Análisis de resultados

```
n=692
x=44
p=.5
p=x/n
sp=((P*(1-P))/n)**.05
# Z estrella
ze=(p-P)/sp
print("Z* =", ze)
# Valor P
vp=norm.cdf(ze)
print("\nValor P =", vp)
Z* = -0.6486481037128928
Valor P = 0.2582829277012232
```

4. Conclusión

Como z^* es menor que z_0 y valor p es mayor que α , entonces, H_1 se rechaza.

Por lo que se puede confirmar que al menos en el lago Tohopekaliga no son dañinos los niveles de mercurio.

Figura 3: Análisis de resultados y conclusión.

Conclusión:

Basándonos en los resultados obtenidos, se puede inferir que las concentraciones de mercurio no presentan un mayor riesgo a la salud de quién llegue a depender de estos o al menos no lo es especialmente en el lago Tohopekaliga. Por lo que se recomendaría aplicar pruebas de hipótesis en los demás lagos para confirmarlo, ya que lógicamente los mismos registros del lago Tohopekaliga no serán los mismos.

Anexos:

Liga al código:

<https://drive.google.com/file/d/1bnLGh5D4suO34Jiz8CwpCSoV4nCdeHUK/view?usp=sharing>