

Reporte final de "Los peces y el mercurio"

Sebastian Rodriguez Salinas - A00827463

Módulo 5: Estadística Avanzada para ciencia de datos y nombre de la concentración.(Gpo 501)



Instrucciones

1. Resumen (espacio máximo: 10 renglones):
 - Introduce la problemática muy brevemente.
 - Indica la forma en que lo abordaste (metodos y técnicas estadísticas)
 - Enuncia tus principales resultados y conclusiones
2. Introducción (espacio máximo: una hoja):
 - Plantea el problema a resolver: puede estar en forma de preguntas.
 - Comenta la importancia del problema. Introduce algunas referencia vinculadas al problema si lo crees necesario.
3. Análisis de los resultados (espacio libre)
 - Estructura tus procedimientos y comenta tus resultados. Usa títulos y subtítulos.
 - Incluye gráficos, tablas o procedimientos, pero sólo los esenciales y añade comentarios o explicaciones a cada uno.
 - No incluyas código en tu reporte.
 - Justifica las herramientas estadísticas usadas y verifica los supuestos de los modelos probabilísticos usados. En el caso de las pruebas de hipótesis justifica la distribución muestral usada.
4. Concluye (espacio máximo: media página).
 - Emite una conclusión general que se vincule con la problemática planteada en la introducción.
5. Referencias bibliográficas: en caso de que hayas consultado una fuente bibliográfica, inclúyela.
6. Anexos:
 - Incluye una liga en web a una carpeta que contenga tus documentos de análisis (R y Python) y el archivo de la base de datos. La liga puede ser a una carpeta en Drive o Google o a un notebook en web.

- Puedes anexar comentarios sobre tus intentos de análisis que no hayan fructificado o que consideres que no deben ir en el reporte, pero que creas importantes.

Resumen

El pescado es uno de los alimentos que mayor valor nutrimental aporta, sin embargo, ¿mucho de este te puede hacer daño? La verdad puede ser más complicada que un simple sí o no. Los pescados son muy señalados en la sociedad por tener concentraciones de mercurio que si pueden llegar a ser dañinos para el humano. Por medio de análisis estadísticos y pruebas de correlación, podemos determinar cómo diferentes variables pueden afectar los niveles de concentración del pescado. Veremos que hay factores que aportan de una manera más directa el nivel de mercurio al igual que cómo se relacionan unos con los otros dando a entender los resultados. Apoyándonos a entender qué factores son los que determinan esta variable que nos puede llegar. afectar en nuestro día a día.

Introducción

El problema a resolver es muy sencillo. Buscamos encontrar los factores que contribuyen a la concentración de mercurio que puede contener los pescados en ciertas áreas. Con esto se analizaron las muestras de diferentes zonas y de esta manera hacer el análisis de los factores contribuyentes a este. Es importante tomar en cuenta ya que el mercurio es un elemento tóxico para la salud de las personas. De acuerdo con una nota publicada en El Periodico en 2022, estudios señalan que una lata de atún tiene unos 15 microgramos de mercurio, mientras que el límite para un adulto es de unos 110 microgramos semanales. Esto nos da a entender que como máximo podemos estar consumiendo 7 latas de atún semanalmente. Dicho esto, el atún es de los pescados más conocidos por tener altas concentraciones de mercurio. Sin embargo, muchos de los animales marinos que son consumidos también tienen altos niveles de mercurio. Esto se debe en parte a los factores externos en los que se sitúan esta vida marina. Tal y como puede ser el ph del agua, la alcalinidad, el calcio, entre otros.

Por esta razón, es de ayuda hacer un análisis que nos ayude a determinar cómo y en qué medida diferentes atributos pueden afectar el mercurio de esto. Teniendo esta se pueden tomar mejores decisiones en el ambiente y en los valores proporcionando un mejor valor nutrimental a los consumidores.

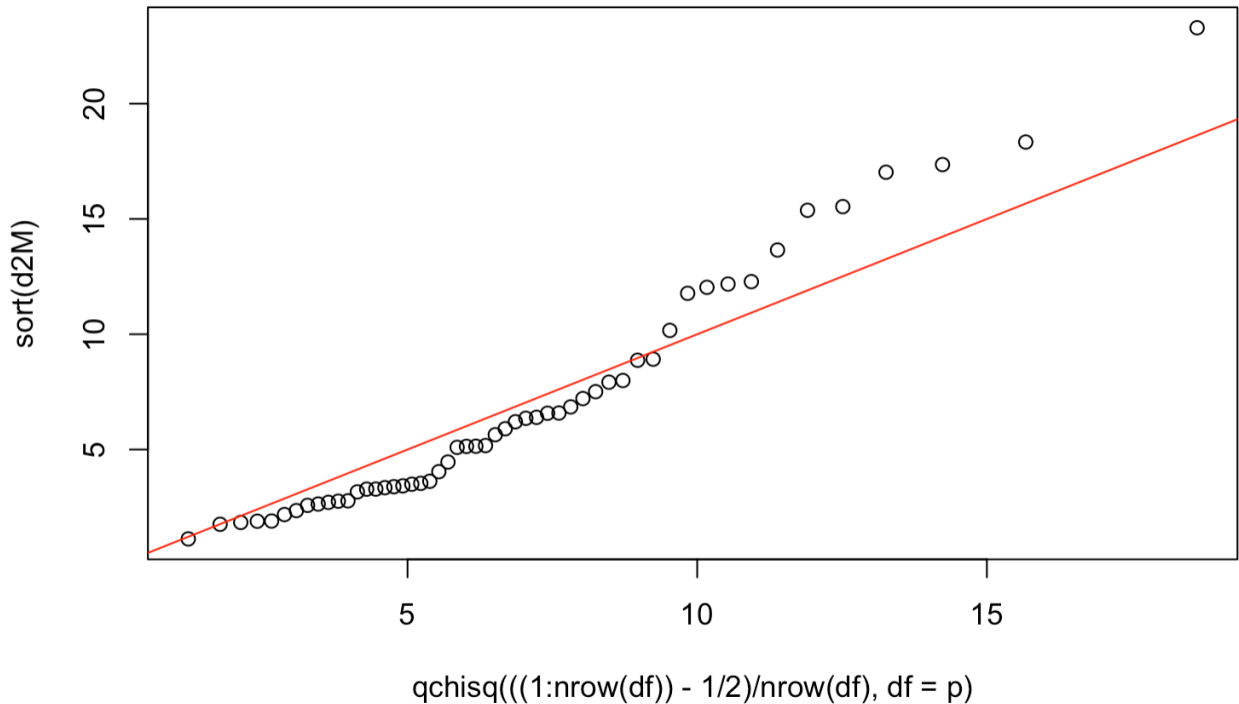
Análisis de los resultados

Lo primero a realizar es la limpieza de datos. Esto dado a que en primera instancia al solo ver los datos de las columnas, podemos ver que las variables de id y name son variables básicas que no nos son de importancia para el análisis de datos. Así mismo, también hay que eliminar los datos de min_mercury, max_mercury y unknown_mercury ya que son valores extras que no nos aportan o incluso nos puede llevar a resultados ambiguos. Estamos dejando la columna de mercury, la cual representa el promedio de mercurio en los peces.

id	name	alkalinity	ph	calcium	chlorophyll	mercury	count	min_mercury	max_mercury	unknown_mercury	age
<int>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<int>
1	Alligator	5.9	6.1	3.0	0.7	1.23	5	0.85	1.43	1.53	1
2	Annie	3.5	5.1	1.9	3.2	1.33	7	0.92	1.90	1.33	0
3	Apopka	116.0	9.1	44.1	128.3	0.04	6	0.04	0.06	0.04	0
4	Blue Cypress	39.4	6.9	16.4	3.5	0.44	12	0.13	0.84	0.44	0
5	Brick	2.5	4.6	2.9	1.8	1.20	12	0.69	1.50	1.33	1

	alkalinity	ph	calcium	chlorophyll	mercury	count	age
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<int>
1	5.9	6.1	3.0	0.7	1.23	5	1
2	3.5	5.1	1.9	3.2	1.33	7	0
3	116.0	9.1	44.1	128.3	0.04	6	0
4	39.4	6.9	16.4	3.5	0.44	12	0
5	2.5	4.6	2.9	1.8	1.20	12	1

Podemos realizar una visualización de datos inicial que nos permite determinar la tendencia y relación de los atributos que estamos por analizar.



Esta gráfica qqplot nos representa la tendencia positiva que hay entre los atributos y el contenido de mercurio en los peces.

Teniendo los datos limpios podemos comenzar el análisis realizando las matrices de covarianza y correlación. estas matrices nos permite despues realizar el análisis de componentes con varianza y correlación, el cual nos da los siguientes resultados:

Varianza

Análisis de componentes con varianza

```
[1] 0.7264306 0.9301116 0.9775598 0.9997295
0.9999440 0.9999800 1.0000000
```

De acuerdo con los resultados, podemos distinguir que los primeros 3 componentes son los más importantes ya que la suma acumulada de las varianzas de los primeros tres atributos representa el 97% de la varianza total. Estos 3 atributos son: alkalinity, chlorophyll y calcium. Estos se pueden

Correlación

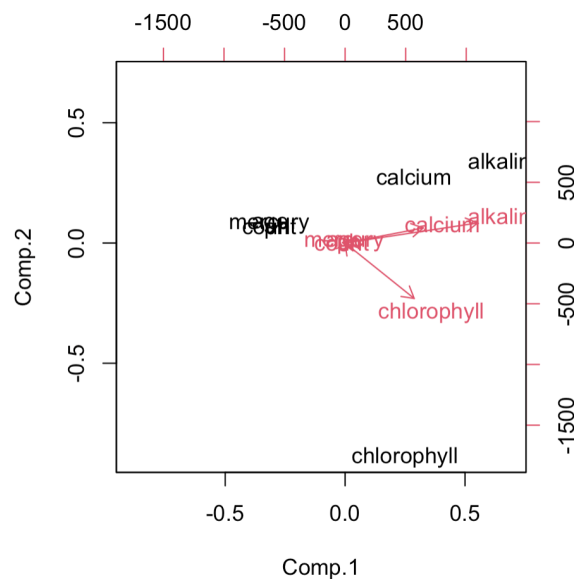
Análisis de componentes con correlación

```
[1] 0.4735349 0.6508486 0.7795647 0.8732057
0.9447597 0.9855166 1.0000000
```

Por el lado de la correlación, podemos distinguir que los primeros 5 componentes son los más importantes ya que la suma acumulada representa el 97% de la varianza total. Estos 5 atributos son: alkalinity, chlorophyll, calcium, ph y mercury. Dejando de lado count y ph. Estos se pueden distinguir

distinguir en la gráfica de vectores representando la varianza. Dicho esto, las unidades de medición para los diferentes atributos difieren, por lo que no se puede hacer una correlación exacta alterando ligeramente los posibles valores de la varianza.

Gráficos de componentes

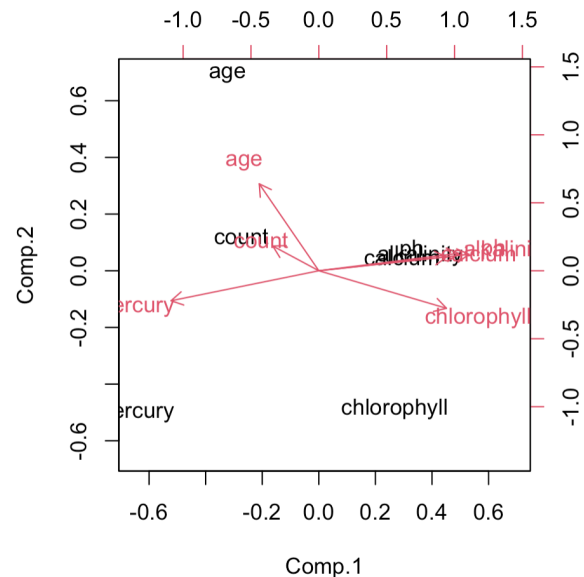


Esta gráfica representa la contribución de parte de alkalinity, chlorophyll y calcium, siendo estos los que representan un mayor vector en este mapa de vectores.

Componentes principales

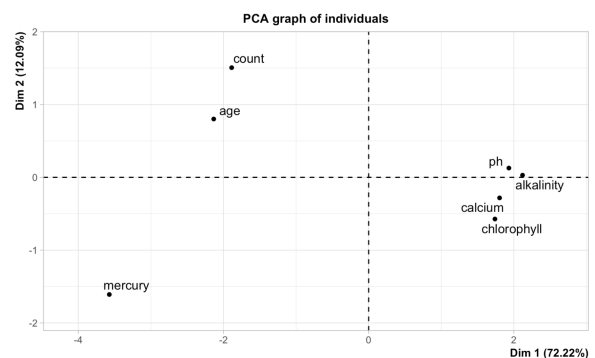
en la gráfica de vectores representando la correlación.

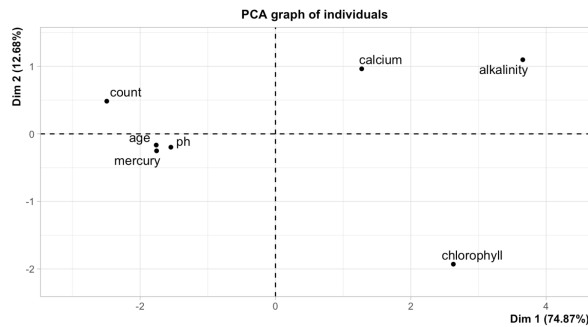
Gráficos de componentes



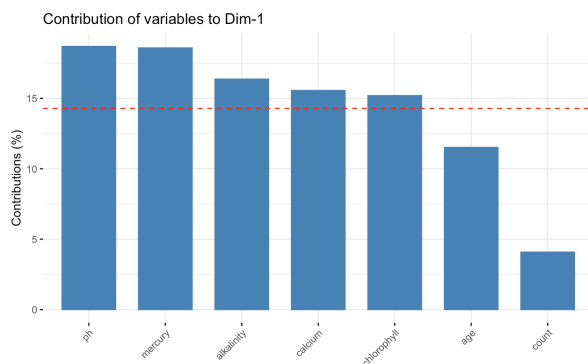
Esta gráfica representa la contribución de parte de alkalinity, chlorophyll, calcium, ph y mercury, siendo estos los que representan un mayor vector en este mapa de vectores.

Componentes principales



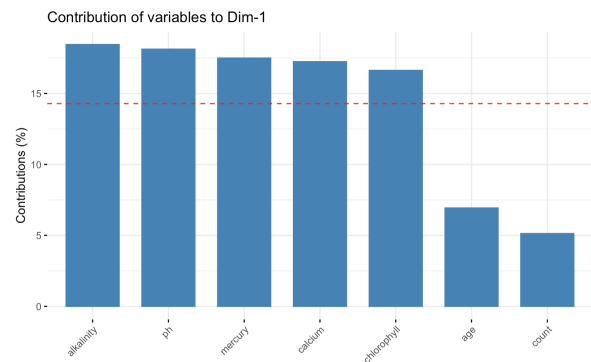


La gráfica 1 de componentes nos representa a los datos alkalinity, chlorophyll y calcium en la dimensión 1, según esto representando el 75% de la predicción de mercurio en los peces.



Este gráfico de barras representa la distribución de las varianzas de cada atributo. Mientras más cercano esté el valor a la línea roja, este valor representa una mayor contribución a los niveles de mercurio del pescado. Dicho esto podemos plantar lo que hemos dicho anteriormente, se representa la contribución de parte de alkalinity, chlorophyll y calcium a la predicción del nivel de mercurio del pescado.

La gráfica 1 de componentes nos representa a los datos alkalinity, chlorophyll, calcium y ph, en la dimensión 1, según esto representando el 72% de la predicción de mercurio en los peces. Así mismo podemos ver que el valor de mercurio está muy retirado lo que no se representa que tiene alguna contribución directa, sin embargo no es determinado proporcionalmente como lo son los otros 4 atributos.



De la misma manera, esta gráfica representa la contribución de parte de alkalinity, chlorophyll, calcium, ph y mercury, a la predicción del nivel de mercurio del pescado. Esto siendo que están relativamente igual de cercanas a la línea roja en comparación a age y count que están muy por debajo, lo que nos da a entender que estos dos contribuyen poco o nada a la predicción de mercurio.

Conclusión

En conclusión, encontramos dentro de los datos de alkalinity, chlorophyll, calcium y ph influyen dentro de los niveles de concentración de mercurio dado que los datos estadísticos indican una gran correlación entre este. Se presenta como en las gráficas se relacionan los atributos a la media de mercurio encontrado en los pescados. Teniendo estos valores se pueden hacer modificaciones al ambiente en el que estos son sometidos con el fin de encontrar mejoras y posibles niveles mercurios más bajos y controlados que puedan contribuir en una menor medida a los altos niveles de mercurio.

Dicho esto, podemos finalizar diciendo que si hay una fuerte relación entre los atributos analizados y los niveles de mercurio encontrados. Tal es la importancia de mantener a los pescados en una atmósfera controlada para su desarrollo.

Referencias

- Modulo 5. Estadística - Blanca Rosa Ruiz Hernandez
- <https://www.elperiodico.com/es/vida-y-estilo/20220705/numero-maximo-latas-atun-comer-mercurio-dv-14008839>

Anexos

- Código R:
https://drive.google.com/file/d/1WI2EqfeH0D0HKRYhixXDe2jJFhHgKO4e/view?usp=share_link
- Dataset: https://drive.google.com/file/d/1M6w_iPJS3pgdrMPcQ-TeEib1nbHoDQ-/view?usp=share_link