

TAREA 12 EJERCICIO DE COMPONENTES PRINCIPALES

La contaminación por mercurio de peces de agua dulce comestibles es una amenaza directa contra nuestra salud. Entre 1990 y 1991 se llevó a cabo un estudio en 53 lagos de Florida con el fin de examinar los factores que influían en el nivel de contaminación por mercurio. Las variables que se midieron fueron:

NUMERO DE IDENTIFICACION
NOMBRE DEL LAGO
X1: ALCALINIDAD (MG CARBONATO DE CALCIO)
X2: PH
X3: CALCIO
X4: CLOROFILA
X5: CONCENTRACION MEDIA DE MERCURIO EN EL TEJIDO MUSCULAR DEL GRUPO DE PECES
X6: NUMERO DE PECES ESTUDIADOS
X7: MINIMA DE CONCENTRACION DE MERCURIO EN CADA GRUPO DE PECES
X8: MAXIMA DE CONCENTRACION DE MERCURIO EN CADA GRUPO DE PECES
X9: ESTIMACION MEDIA DE CONCENTRACION DE MERCURIO EN PECES DE 3 AÑOS
X10: INDICADOR DE LA EDAD DE LOS PECES

REALIZA EL ANALISIS DE COMPONENTES PRINCIPALES UTILIZANDO LA MATRIZ DE CORRELACION. SELECCIONA EL NUMERO DE COMPONENTES PRINCIPALES CUYO VALOR PROPIO SE MAYOR O IGUAL QUE UNO. EN CADA COMPONENTE SELECCIONA LAS VARIABLES CUYAS CARGAS SEAN LAS MAS GRANDES.

No. identificació n	Nombre del lago	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
1	Alligator	5.9	6.1	3	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	1.9	1.33	0
3	Apopka	116	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	1.2	12	0.69	1.5	1.33	1
6	Bryant	19.6	7.3	4.5	44.1	0.27	14	0.04	0.48	0.25	1
7	Cherry	5.2	5.4	2.8	3.4	0.48	10	0.3	0.72	0.45	1
8	Crescent	71.4	8.1	55.2	33.7	0.19	12	0.08	0.38	0.16	1
9	Deer Point	26.4	5.8	9.2	1.6	0.83	24	0.26	1.4	0.72	1
10	Dias	4.8	6.4	4.6	22.5	0.81	12	0.41	1.47	0.81	1
11	Dorr	6.6	5.4	2.7	14.9	0.71	12	0.52	0.86	0.71	1
12	Down	16.5	7.2	13.8	4	0.5	12	0.1	0.73	0.51	1
13	Eaton	25.4	7.2	25.2	11.6	0.49	7	0.26	1.01	0.54	1
14	East Tohopekaliga	7.1	5.8	5.2	5.8	1.16	43	0.5	2.03	1	1
15	Farm- 13	128	7.6	86.5	71.1	0.05	11	0.04	0.11	0.05	0
16	George	83.7	8.2	66.5	78.6	0.15	10	0.12	0.18	0.15	1
17	Griffin	108.5	8.7	35.6	80.1	0.19	40	0.07	0.43	0.19	1
18	Harney	61.3	7.8	57.4	13.9	0.77	6	0.32	1.5	0.49	1
19	Hart	6.4	5.8	4	4.6	1.08	10	0.64	1.33	1.02	1
20	Hatchineha	31	6.7	15	17	0.98	6	0.67	1.44	0.7	1
21	Iamonia	7.5	4.4	2	9.6	0.63	12	0.33	0.93	0.45	1
22	Istokpoga	17.3	6.7	10.7	9.5	0.56	12	0.37	0.94	0.59	1
23	Jackson	12.6	6.1	3.7	21	0.41	12	0.25	0.61	0.41	0
24	Josephine	7	6.9	6.3	32.1	0.73	12	0.33	2.04	0.81	1
25	Kinglsey	10.5	5.5	6.3	1.6	0.34	10	0.25	0.62	0.42	1
26	Kissimmee	30	6.9	13.9	21.5	0.59	36	0.23	1.12	0.53	1
27	Lochloosa	55.4	7.3	15.9	24.7	0.34	10	0.17	0.52	0.31	1
28	Lousia	3.9	4.5	3.3	7	0.84	8	0.59	1.38	0.87	1
29	Miccasukee	5.5	4.8	1.7	14.8	0.5	11	0.31	0.84	0.5	0
30	Minneola	6.3	5.8	3.3	0.7	0.34	10	0.19	0.69	0.47	1
31	Monroe	67	7.8	58.6	43.8	0.28	10	0.16	0.59	0.25	1
32	Newmans	28.8	7.4	10.2	32.7	0.34	10	0.16	0.65	0.41	1
33	Ocean Pond	5.8	3.6	1.6	3.2	0.87	12	0.31	1.9	0.87	0

34	Ocheese Pond	4.5	4.4	1.1	3.2	0.56	13	0.25	1.02	0.56	0
35	Okeechobee	119.1	7.9	38.4	16.1	0.17	12	0.07	0.3	0.16	1
36	Orange	25.4	7.1	8.8	45.2	0.18	13	0.09	0.29	0.16	1
37	Panasoffkee	106.5	6.8	90.7	16.5	0.19	13	0.05	0.37	0.23	1
38	Parker	53	8.4	45.6	152.4	0.04	4	0.04	0.06	0.04	0
39	Placid	8.5	7	2.5	12.8	0.49	12	0.31	0.63	0.56	1
40	Puzzle	87.6	7.5	85.5	20.1	0.1	10	0.79	1.41	0.89	1
41	Rodman	114	7	72.6	6.4	0.16	14	0.04	0.26	0.18	1
42	Rousseau	97.5	6.8	45.5	6.2	0.1	12	0.05	0.26	0.19	1
43	Sampson	11.8	5.9	24.2	1.6	0.48	10	0.27	1.05	0.44	1
44	Shipp	66.5	8.3	26	68.2	0.21	12	0.05	0.48	0.16	1
45	Talquin	16	6.7	41.2	24.1	0.86	12	0.36	1.4	0.67	1
46	Tarpon	5	6.2	23.6	9.6	0.52	12	0.31	0.95	0.55	1
47	Trafford	81.5	8.9	20.5	9.6	0.27	6	0.04	0.4	0.27	0
48	Trout	1.2	4.3	2.1	6.4	0.94	10	0.59	1.24	0.98	1
49	Tsala Apopka	34	7	13.1	4.6	0.4	12	0.08	0.9	0.31	1
50	Weir	15.5	6.9	5.2	16.5	0.43	11	0.23	0.69	0.43	1
51	Tohopekaliga	25.6	6.2	12.6	27.7	0.65	44	0.3	1.1	0.58	1
52	Wildcat	17.3	5.2	3	2.6	0.25	12	0.15	0.4	0.28	1
53	Yale	71.8	7.9	20.5	8.8	0.27	12	0.15	0.51	0.25	1

Análisis de componentes principales para la contaminación de los lagos

Para aplicar el Análisis de Componentes Principales (CP), de los 53 lagos con las 10 variables de contaminación, se utilizó la matriz de correlación. Considerando el criterio del valor propio mayor que uno, se puede ver en la Tabla 1 y en la figura 1, que son 3 componentes principales los cuales cumplen con el criterio. Y también se puede ver que con 3 componentes principales se explica el 77.30% de la variabilidad total de los datos. Donde el primer componente explica el 53.68% de la variabilidad total, el segundo componente, explica el 12.49% de la variabilidad total de la información, y el tercer componente principal explica el 11.11% de la variación total. Lo que implica que, con solo 3 componentes principales, se puede analizar e interpretar los resultados de la contaminación de los lagos.

<i>Component Number</i>	<i>Eigenvalue</i>	<i>Percent of Variance</i>	<i>Cumulative Percentage</i>
1	5.36866	53.687	53.687
2	1.24987	12.499	66.185
3	1.11146	11.115	77.300
4	0.918752	9.188	86.487
5	0.59511	5.951	92.439
6	0.307567	3.076	95.514
7	0.20705	2.070	97.585
8	0.118368	1.184	98.768
9	0.0770467	0.770	99.539
10	0.0461172	0.461	100.000

Tabla 1. Porcentaje de varianza de los componentes principales

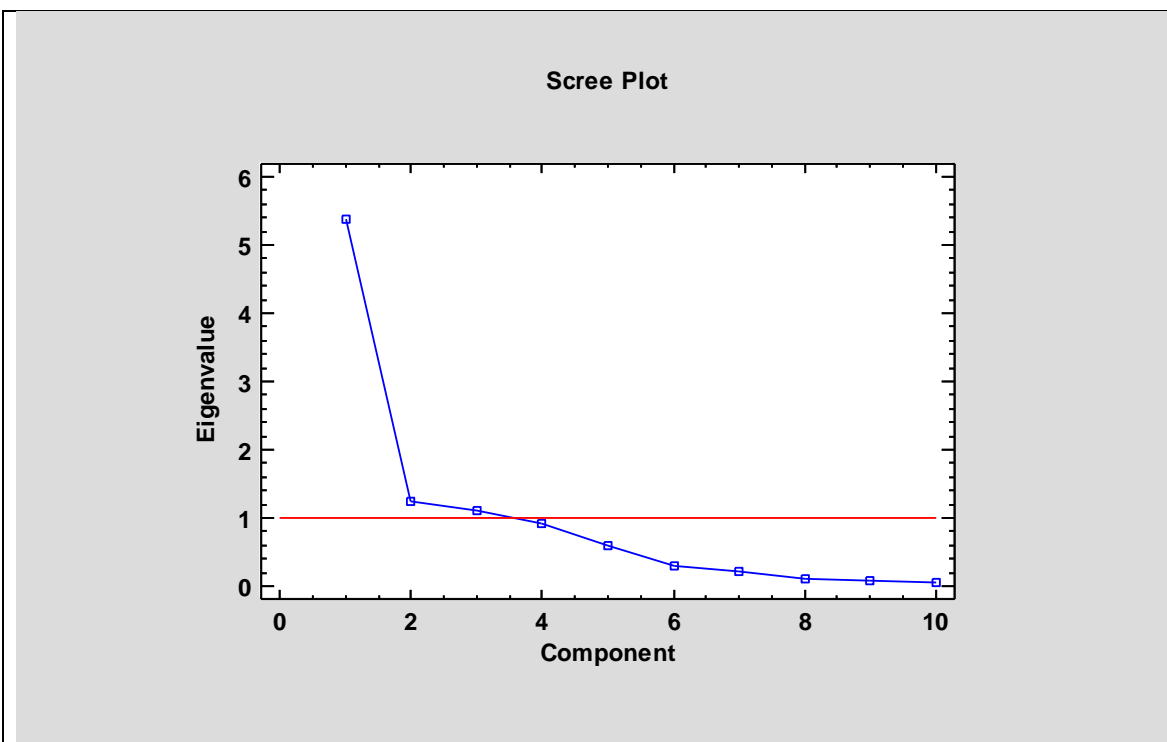


Figura 1. Grafica de sedimentación

Para identificar la naturaleza de estos 3 componentes principales es necesario analizar las variables de contaminación que se asocian a cada componente principal, lo cual se puede observar en la matriz de cargas o pesos, que en este caso se muestra en tabla 2. Una variable se relaciona principalmente con aquel componente principal donde su carga o peso sea claramente mayor que las cargas o pesos de las demás variables dentro del mismo componente. En la tabla 2 se pueden observar todas las variables de contaminación seleccionadas mediante el criterio de las cargas más grande, tomando el criterio de considerar las cargas mayores o iguales que a 0.25, en valor absoluto. Bajo este

criterio se puede ver que el primer componente principal se asocia claramente a 8 variables contaminantes; en el segundo componente principal se relacionan 2 variables contaminantes y en el tercer componente se explican 4 variables de contaminación.

	<i>Component 1</i>	<i>Component 2</i>	<i>Component 3</i>
(X1: ALCALINIDAD (MG CARBONATO DE CALCIO))	0.358113	0.100063	0.38312
(X2: PH)	0.341187	0.134975	0.272648
(X3: CALCIO)	0.295132	0.0962706	0.569026
(X4: CLOROFILA)	0.283327	-0.180693	0.197572
(X5: CONCENTRACION MEDIA DE MERCURIO EN EL TEJIDO MUSCULAR DEL GR)	-0.398558	0.00831036	0.19021
(X6: NUMERO DE PECES ESTUDIADOS)	-0.0287341	0.638099	-0.139127
(X7: MINIMA DE CONCENTRACION DE MERCURIO EN CADA GRUPO DE PECES)	-0.357578	-0.0684872	0.431561
(X8: MAXIMA DE CONCENTRACION DE MERCURIO EN CADA GRUPO DE PECES)	-0.374029	0.0796773	0.278916
(X9: ESTIMACION MEDIA DE CONCENTRACION DE MERCURIO EN PECES DE 3)	-0.396647	-0.021034	0.310467
(X10: INDICADOR DE LA EDAD DE LOS PECES)	-0.0566297	0.714933	0.0095703
Tabla 2. Cargas o pesos de los componentes principales			

En la tabla 3 se pueden ver las 8 variables de contaminación que son explicadas por el primer componente principal. Estas variables corresponden a la contaminación de la Alcalinidad, pH, Calcio y Clorofila con valores de cargas positivas y las variables de contaminación referentes a la mínima, media, máxima y estimación media de mercurio con valores de cargas negativas. Por el comportamiento de estas variables en este componente principal se puede decir que hay lagos que se caracterizan por una contaminación de altos valores de

Alcalinidad, pH, Calcio y Clorofila y valores bajos de las variables de mínimo, media, máxima y estimación media de mercurio, o viceversa.

COMPONENTE PRINCIPAL 1	<i>cargas</i>
(X1: ALCALINIDAD (MG CARBONATO DE CALCIO))	0.358113
(X2: PH)	0.341187
(X3: CALCIO)	0.295132
(X4: CLOROFILA)	0.283327
(X5: CONCENTRACION MEDIA DE MERCURIO EN EL TEJIDO MUSCULAR DEL GR)	-0.398558
(X7: MINIMA DE CONCENTRACION DE MERCURIO EN CADA GRUPO DE PECES)	-0.357578
(X8: MAXIMA DE CONCENTRACION DE MERCURIO EN CADA GRUPO DE PECES)	-0.374029
(X9: ESTIMACION MEDIA DE CONCENTRACION DE MERCURIO EN PECES DE 3)	-0.396647

Tabla 3. Variables explicadas por el primer componente principal

En la tabla 4 están las 2 variables que son explicadas por el segundo componente principal. Estas variables corresponden al número estudiado de peces y al indicador de la edad de los peces. Este componente indica hay lagos que se caracterizan por la cantidad de peces estudiados y por el indicador de la edad del pez.

COMPONENTE PRINCIPAL 1	<i>Cargas</i>
(X6: NUMERO DE PECES ESTUDIADOS)	0.638099
(X10: INDICADOR DE LA EDAD DE LOS PECES)	0.714933

Tabla 4. Variables explicadas por el segundo componente principal

En el tercer componente principal explica a 4 variables de contaminación (ver tabla 5). Estas 4 variables son la Alcalinidad, Calcio, mínimo y estimación media de mercurio, lo que significa que hay lagos que se caracterizan por valores altos tanto de Alcalinidad, Calcio y valores altos de concentración mínima y estimación media de mercurio.

COMPONENTE PRINCIPAL 3	<i>cargas</i>
(X1: ALCALINIDAD (MG CARBONATO DE CALCIO))	0.38312
(X3: CALCIO)	0.569026
(X7: MINIMA DE CONCENTRACION DE MERCURIO EN CADA GRUPO DE PECES)	0.431561
(X9: ESTIMACION MEDIA DE CONCENTRACION DE MERCURIO EN PECES DE 3)	0.310467

Tabla 5. Variables explicadas por el tercer componente principal

Grafica scatterplot

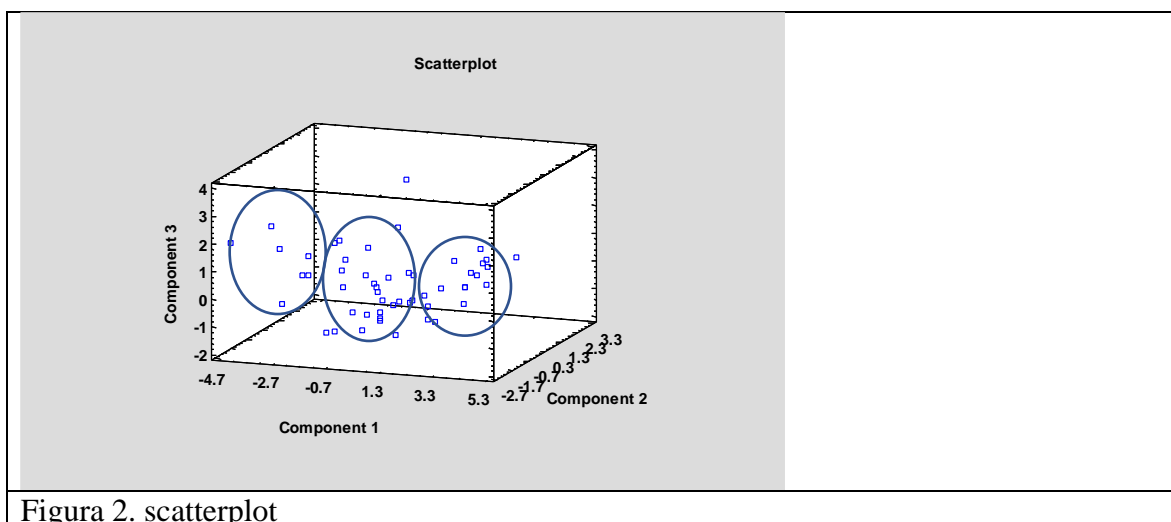


Figura 2. scatterplot

En la figura 2 se puede observar los tres tipos de lagos caracterizados por la contaminación según los componentes principales.

GRAFICA DE LOS PESOS DE LOS COMPONENTES PRINCIPALES

En la figura 3, se puede ver el grafico de los pesos de los componentes, obsérvese el ajuste y la explicación de las variables de contaminación en los componentes 1 y 2.

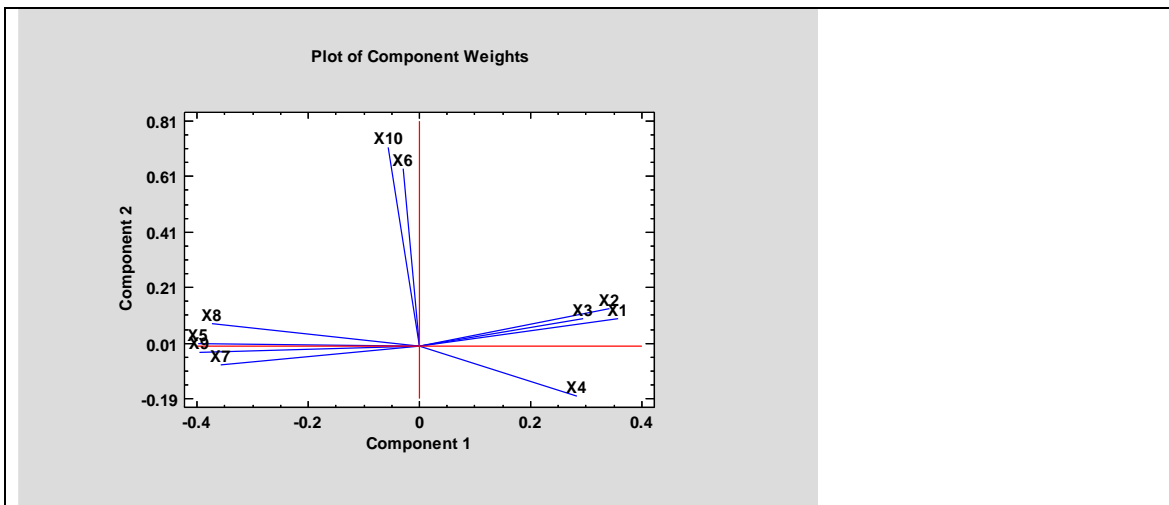


Figura 3. Grafica de los pesos de los componentes principales