

# Una aplicación de Componentes Principales: Tamaño y Forma del Caparazón de Tortuga Pintada

*Jaime Andrés Gaviria Bedoya*<sup>a</sup>, *César Augusto Flórez Restrepo*<sup>b</sup>, *Sergio Yañez Canal*<sup>c</sup>

Email: jagaviriab@unal.edu.co

a. *Estudiante maestría en Estadística, Universidad Nacional de Colombia-Sede Medellín*

b. *Estudiante maestría en Estadística, Universidad Nacional de Colombia-Sede Medellín*

c. *Profesor Asociado, Universidad Nacional de Colombia-Sede Medellín*

## Resumen

El problema de explicar fenómenos naturales y sociales se ha convertido en una necesidad fundamental para la ciencia. Uno de los problemas importantes en Biología es modelar el crecimiento de los organismos vivos, mediante la ecuación alométrica. Se ilustra una aplicación del Análisis de Componentes Principales de las medidas del caparazón de las tortugas pintadas de Canadá, para determinar el tamaño y forma en función de la longitud, el alto y el ancho. Se muestra como las componentes principales dan los elementos para generalizar la ecuación alométrica (ecuación de crecimiento) a varias variables, utilizando la transformación logarítmica. Se trata de un artículo divulgativo que ilustra la importancia del análisis multivariado para solución de problemas biológicos.

# A Principal Component Analysis: Size and Shape Variation in the Painted Turtle

*Jaime Andrés Gaviria Bedoya*<sup>a</sup>, *César Augusto Flórez Restrepo*<sup>b</sup>, *Sergio Yañez Canal*<sup>c</sup>

Email: jagaviriab@unal.edu.co

---

- a. *Masters student in statistics, Universidad Nacional de Colombia-Sede Medellín*
- b. *Masters student in statistics, Universidad Nacional de Colombia-Sede Medellín*
- c. *Associate Professor, Universidad Nacional de Colombia-Sede Medellín*

### **Abstract**

The problem of explaining natural and social phenomena has become a fundamental requirement for science. One of the important problems in biology is to model the growth of living organisms, using the allometric equation. We illustrate an application of Principal Component Analysis of the measurements of painted turtle in Canada to determine the size and shape depending on the length, height and width. It shows how the main components are the elements to generalize the allometric equation (equation of growth) for several variables, using the logarithmic transformation. This is an informative article that illustrates the importance of multivariate analysis for troubleshooting Biological relevance.

## 0.1. Introducción

El problema de explicar fenómenos naturales y sociales se ha convertido en una necesidad fundamental para la ciencia. En estadística, existen distintos métodos que permiten una mejor descripción de los datos. Uno de ellos es el Análisis de Componentes Principales, que en el área de la biología ha mostrado ser bastante útil. Uno de los problemas fundamentales en esta área es modelar el crecimiento de los organismos, el cual está muy relacionado con el tamaño y la forma del individuo. Una dificultad que surge es lograr cuantificar estas características.

Una solución a este problema fue la propuesta por Pierre Jolicoeur y James Mosimann en 1960 (1). Ellos tomaron una muestra representativa de 24 tortugas (*Chrysemys Picta Marginada*) macho y hembra de todas las edades en el sur de Montreal (Canadá), y midieron la longitud, el ancho y el altura del caparazón. Después de un análisis exploratorio, realizaron un Análisis de Componentes Principales e interpretaron la variación del tamaño y forma del caparazón en la dirección del incremento del volumen. La matriz de covarianzas la asumieron igual en todas las poblaciones de tortugas, lo que es inadecuado si tenemos en cuenta las condiciones climáticas en las cuales se encuentran. Por ello en 1963 Pierre Jolicoeur utiliza la ecuación alométrica de crecimiento, la cual considera unas cantidades a estimar que dependen de las condiciones de vida de estos organismos, estas cantidades determinan el volumen del caparazón.

Inicialmente otros autores habían planteado un análisis de regresión lineal múltiple, procedimiento inadecuado al ignorar la correlación o multicolinealidad de las variables predictoras. Jolicoeur aplica el método de componentes principales y sugiere posteriormente la transformación mediante el logaritmo natural. Implícitamente considera la multicolinealidad y la relación alométrica entre la variable respuesta con las variables predictoras de los datos y estima el volumen del caparazón para la población de machos y hembras de esta región en particular. Muestra los estimadores y el volumen del caparazón, resultado de la primera componente principal, la cual explica un porcentaje de variabilidad muy alto en comparación con las demás componentes principales.

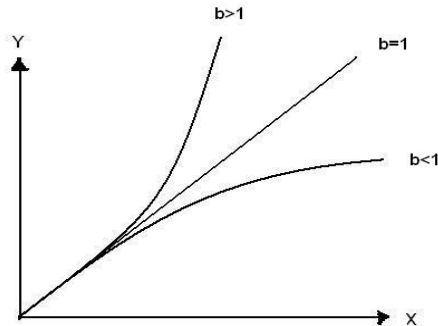
## 0.2. Métodos

En el estudio del crecimiento de los organismos, se debe tener en cuenta la Ecuación Alométrica, la cual modela el crecimiento de los órganos, mediante la ecuación:

$$y = ax^b, \quad a > 0, \quad b > 0$$

donde  $a$  y  $b$  son cantidades a estimar y dependen de las condiciones climáticas y geográficas en las que se encuentran los organismos.

Gráficamente para  $a$  y  $b$  constantes:



Una desventaja del modelo anterior es que es univariado, sólo permite describir el crecimiento en una dimensión. Una generalización multivariada de la ecuación alométrica está dada por Jolicoeur (2)

$$V = al^{\alpha_1}w^{\alpha_2}h^{\alpha_3}$$

la cual permite modelar el crecimiento de un organismo mediante tres dimensiones:  $l$  longitud,  $w$  ancho, y  $h$  altura.

Jolicoeur (2) estima  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$  en la ecuación del volumen dada por

$$V = l^{\alpha_1}w^{\alpha_2}h^{\alpha_3}$$

la forma de hacerlo es aplicando logaritmo natural

$$\ln(V) = \alpha_1\ln(l) + \alpha_2\ln(w) + \alpha_3\ln(h),$$

este método considera la multicolinealidad de las variables y la relación alométrica del volumen con las variables restantes.

### 0.3. Ejemplo

En biología y mas particularmente en el estudio del crecimiento de los organismos hay dos factores que son de importancia: el tamaño y la forma. Una dificultad que se presenta es poder cuantificar de alguna forma estos factores. Se ilustrará como mediante el análisis de componentes principales, se pueden identificar las componentes principales con el tamaño y la forma y se mostrará la relación de dicha solución con la generalización multivariada de la ecuación alométrica.

En un estudio realizado por Jolicoeur & Mossimann (1), se tomaron 84 especimenes de tortugas (de la especie *Chrysemys Picta Marginada*) el 2 de agosto de 1956 a 35 millas del sur de Montreal, Canadá. Luego seleccionaron una muestra representativa de 24 machos y hembras y a cada caparazón le midieron la Longitud, el Ancho y el Altura. El objetivo del estudio era poder modelar el crecimiento del caparazón de las tortugas identificando aquellos factores que describian la variación en el tamaño y la forma del caparazón

Los datos se dan a continuación:

TABLE 1  
 CARAPACE DIMENSIONS OF PAINTED TURTLES (*Chrysemys picta marginata*) IN MM.

24 Males			24 Females		
length	width	height	length	width	height
93	74	37	98	81	38
94	78	35	103	84	38
96	80	35	103	86	42
101	84	39	105	86	40
102	85	38	109	88	44
103	81	37	123	92	50
104	83	39	123	95	46
106	83	39	133	99	51
107	82	38	133	102	51
112	89	40	133	102	51
113	88	40	134	100	48
114	86	40	136	102	49
116	90	43	137	98	51
117	90	41	138	99	51
117	91	41	141	105	53
119	93	41	147	108	57
120	89	40	149	107	55
120	93	44	153	107	56
121	95	42	155	115	63
125	93	45	155	117	60
127	96	45	158	115	62
128	95	45	159	118	63
131	95	46	162	124	61
135	106	47	177	132	67

Al inspeccionar los datos se observa que existen diferentes fuentes de variación:

- Variación en tamaño: Refleja la relación de proporcionalidad entre longitud, ancho y altura, en el sentido de que las tres medidas tienden a aumentar o disminuir simultáneamente.
- Variación en forma: Refleja las diferencias en el ancho y la altura del caparazón para tortugas cuya longitud es la misma.
- Variación en sexo: Las mediciones del ancho, largo y alto en los caparazones de las hembras son mayores que las de los machos.

Dado que en los estudios sobre el crecimiento es común trabajar con el logaritmo de las variables, los siguientes análisis se aplican al logaritmo de los datos, ver (2).

### 0.3.1. Análisis descriptivo:

En primer lugar se realizó un análisis descriptivo de los datos transformados, discriminando los dos grupos: machos y hembras. Los resultados se presentan a continuación:

Para las tortugas hembra, el vector de medias es :

Variable	LogLongitud	LogAncho	LogAltura
Media	4.900659	4.622909	3.940286

Y la matriz de covarianzas

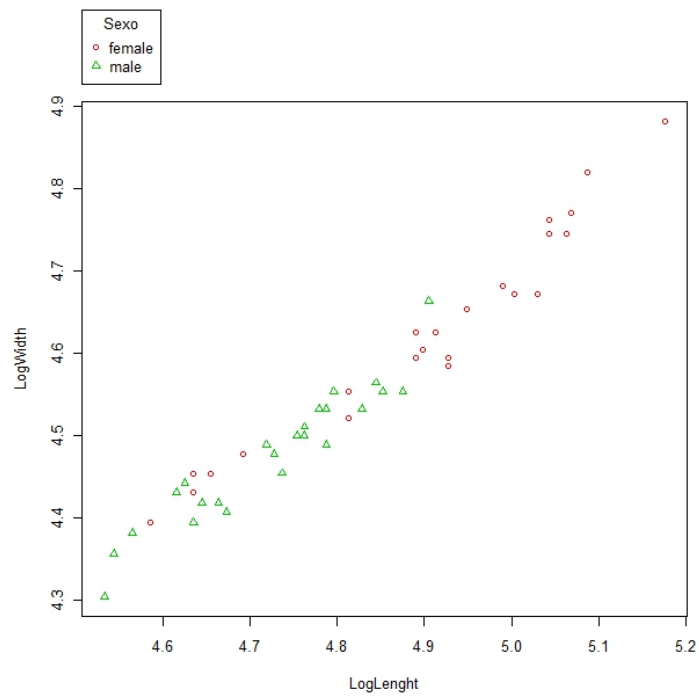
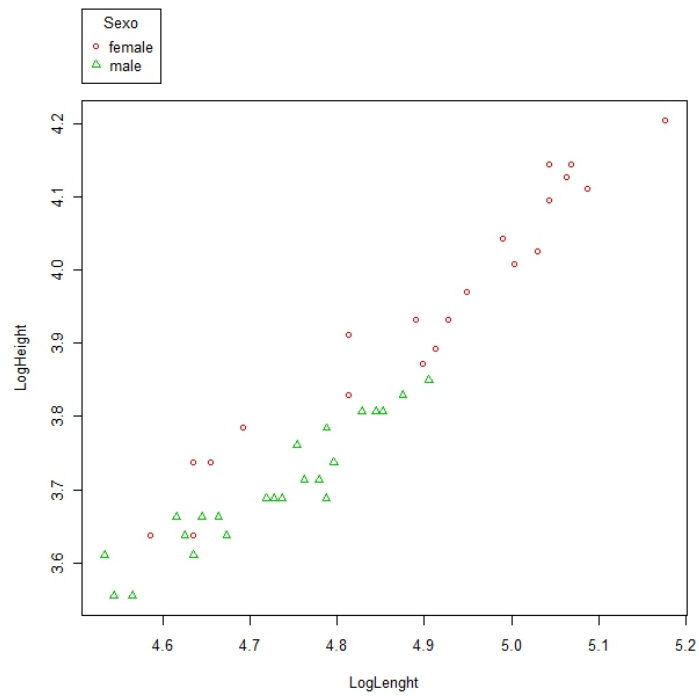
$$S = \begin{bmatrix} 0,02640563 & 0,02011195 & 0,02491758 \\ 0,02011195 & 0,01619045 & 0,01942430 \\ 0,02491758 & 0,01942430 & 0,02493980 \end{bmatrix}$$

Para los machos, el vector de medias y la matriz de covarianzas respectivamente son:

Variable	LogLongitud	LogAncho	LogAltura
Media	4.725	4.478	3.703

$$S = 10^{-3} \begin{bmatrix} 11,072 & 8,019 & 8,160 \\ 8,019 & 6,417 & 6,005 \\ 8,160 & 6,005 & 6,773 \end{bmatrix}$$

Se observa que con la transformación se redujo la diferencia en las mediciones entre machos y hembras, aunque las entradas del vector de medias y la matriz de covarianzas para las hembras son ligeramente mayores que para los machos. Además comparando las varianzas para machos y hembras, se observa que estas últimas presentan una mayor dispersión que los primeros, lo cual se observa en los siguientes gráficos:



### 0.3.2. Análisis de Componentes Principales

Se presenta a continuación el resumen del Análisis de componentes principales realizado a la matriz de covarianzas de las hembras y los machos :

### Análisis para las hembras

Componentes Principales			
Variable	$\hat{e}_1$	$\hat{e}_2$	$\hat{e}_3$
Ln(longitud)	-0,626	-0,552	0,5495
Ln(ancho)	-0,487	-0,271	-0,829
Ln(altura)	-0,607	0,787	0,0999
varianza $\hat{\lambda}_i$	0,0634	0,0007	0,0005
Porcentaje total de Varianza	98,06	99,19	100

La primera componente principal se expresa como:

$$\hat{y}_1 = 0,626\ln(l) + 0,487\ln(w) + 0,607\ln(h)$$

y explica mayormente la variabilidad total con un 98%. Esta componente le da aproximadamente igual peso a las tres mediciones, lo cual indica que una tortuga hembra con puntuaciones altas en esta componente, tendrá también unas mediciones grandes en las tres variables. Esta componente se puede interpretar como un factor de tamaño, ya que consiste en una variación conjunta en las tres dimensiones. Las otras dos componentes aportan muy poco a la variabilidad total, menos del 2% entre las dos. La segunda componente hace un contraste del logaritmo de la longitud del caparazón frente al logaritmo de la altura. La tercera componente contrasta la longitud frente al ancho del caparazón (en logaritmo).

### Análisis para los machos

Componentes Principales			
Variable	$\hat{e}_1$	$\hat{e}_2$	$\hat{e}_3$
Ln(longitud)	0,683	-0,159	-0,713
Ln(ancho)	0,510	-0,594	0,622
Ln(altura)	0,523	0,788	0,324
varianza $\hat{\lambda}_i$	$23,30 \times 10^{-3}$	$0,6 \times 10^{-3}$	$0,36 \times 10^{-3}$
Porcentaje total de Varianza	96,1	98,5	100

La primera componente principal se expresa como:

$$\hat{y}_1 = 0,683\ln(l) + 0,510\ln(w) + 0,523\ln(h)$$

y explica un 96,1% de la variabilidad total. Al igual que para las hembras, esta componente se interpreta como un factor de tamaño. Nuevamente las otras 2 componentes son poco significativas, ya que explican conjuntamente menos del 4% de la variabilidad total. La segunda componente hace un contraste de altura frente al ancho del caparazón, mientras que la tercera componente hace un contraste de la longitud frente al ancho y la altura del caparazón.

## 0.4. Discusión

Esta aplicación del análisis de componentes principales, logra identificar las componentes con dos fuentes de variación: una de tamaño y otra de forma. La primera componente, se interpreta como un factor que explica la variación en el tamaño del caparazón, ya que corresponde a una variación conjunta de las 3 dimensiones. La segunda y tercera componente se interpretan como un factor que explica la variación en la forma del caparazón, ya que hace un contraste en las tres mediciones.

Se observa también que las tortugas hembra presentan una mayor variabilidad en el tamaño (primera componente) que los machos, lo cual concuerda con el análisis descriptivo realizado anteriormente.

Tanto en las hembras como en los machos, la primera componente es una medida de la variación del tamaño del caparazón. Mas aún se puede dar una interpretación de este factor en relación al volumen del caparazón, ver (3). Por ejemplo para los machos, la primera componente principal

$$\hat{y}_1 = 0,683\ln(l) + 0,510\ln(w) + 0,523\ln(h)$$

se puede escribir como

$$y_1 = \ln[l^{0,683}w^{0,510}h^{0,523}]$$

que es exactamente la ecuación alométrica para el logaritmo del volumen del caparazón de las tortugas, donde los coeficientes  $\alpha_1, \alpha_2, \alpha_3$  se han estimado mediante las coordenadas de la primera componente principal, 0,683, 0,510, 0,523, para la longitud, el ancho y la altura del caparazón respectivamente. Una interpretación análoga se observa para la primera componente de las hembras.

## 0.5. Conclusiones

Mediante el estudio descrito, los autores han mostrado como se pueden identificar el tamaño y la forma del caparazón de tortugas, con las componentes principales de la matriz de covarianzas para machos y hembras. Encontraron que la primera componente es la mas significativa y mediante la generalización multivariada de la ecuación alométrica, lograron interpretar dicha componente en términos del logaritmo del volumen, donde los coeficientes son estimados por las coordenadas de la primera componente principal.

# Referencias

- [1] P. Jolicoeur and J. E. Mosimann.(1960). Size and Shape Variation in the Painted Turtle. A Principal Component Analysis. *Growth* 24 :339-354.
- [2] P. Jolicoeur.(1963). The Multivariate Generalization of the Allometry Equation. *Biometrics* 19 :497-499.
- [3] R.A.Johnson and D.W. Wichern.(1998). *Appied Multivariate Statistical Analysis*. New Jersie: Prentice Hall, Inc-4th ed.