

# Analisis de componentes principales

---

Luis Fernando Delgado Muñoz - Ing.  
Agroindustrial, M.Sc

[lfdelgadam@unal.edu.co](mailto:lfdelgadam@unal.edu.co)

Universidad Nacional de Colombia Facultad  
de Ingeniería y Administración Departamento de  
Ingenieria



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

## Análisis de componentes principales

---

El análisis de componentes principales es una técnica de la reducción de la dimensión que describe la información de un conjunto de variables observadas mediante un conjunto de variables mas pequeño (componentes principales), que son combinaciones lineales de las variables de partida.

El objetivo principal del ACP, es comparar individuos según valores de las variables continuas que se parecen.

Por ejemplo dos individuos se parecen porque obtienen más o menos los mismos indicadores en diferentes aspectos.

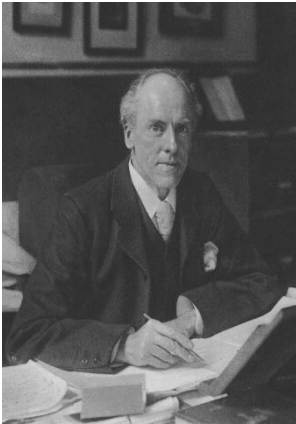
## Objetivos

---

- Generar nuevas variables que pueden expresar la información contenida en el conjunto original de datos.
- Reducir la dimensionalidad del problema que se está estudiando, como paso previo para futuros análisis.
- Eliminar cuando sea posible, algunas de las variables originales si ellas aportan poca información.

# Orígenes del Análisis en Componentes Principales

---



En 1901 Karl Pearson publicó un trabajo sobre el ajuste de un sistema de puntos en un multiespacio a una línea o un plano. Este enfoque fue retomado en 1933 por Hotelling, quien fue el primero en formular el análisis por componentes tal como se ha difundido hasta nuestros días.

El trabajo original de Pearson (1901) se centraba en aquellos componentes, o combinaciones lineales de variables originales, para los cuales la varianza no explicada fuera mínima. Estas combinaciones generaban un plano, función de las variables, en el cual el ajuste del sistema de puntos es “el mejor”, por ser mínima la suma de las distancias de cada punto al plano de ajuste.

## Fundamentos del método

---

El ACP recurre a dos representaciones geométricas: una para comparar a los individuos y otra para estudiar las relaciones entre las variables.

La representación de los datos depende de las unidades y escalas usadas, por lo tanto es necesario que las variables estén expresadas en la misma unidad de medida.

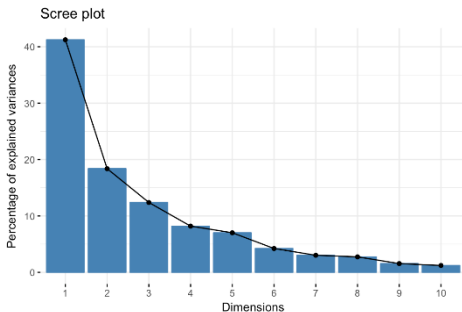
Para ello es conveniente estandarizar los caracteres a media 0 y varianza 1

Estas representaciones requieren de transformaciones de la tabla de datos. La transformación más utilizada es la de la estandarización de los datos, es decir restar la media y dividir por la desviación estándar.

Cada observación tendrá un nuevo valor definido en función de unidades de desvío estándar.

## Fundamentos del método

Los componentes no están correlacionados entre sí y se interpretan independientemente, cada uno capturando parte de la variabilidad total de la matriz básica de datos (MBD). El primer componente (PC1) contiene la mayor variabilidad, seguido por el segundo (PC2) y el tercero (PC3), cada uno con la mayor variabilidad no capturada por los anteriores. Este proceso continúa hasta distribuir toda la variabilidad entre los componentes, cada uno con información de todas las variables en diferentes proporciones.



# Terminología estándar del análisis de componentes principales

---

**Componente principal:** variable hipotética que se construye a partir de las variables originales, no se encuentra correlacionada con otros componentes principales.

**Eigenvalor:** varianza explicada por un componente principal.

**Eigenvector:** vector que define la dirección de un componente principal.

**Loading:** se define como  $\text{loading} = \text{eigenvector} * \sqrt{\text{eigenvalor}}$

**Círculo de correlación:** gráfico en el cual se representan las variables junto con una circunferencia de radio igual a 1 y que permite visualizar la calidad de la representación de las variables.

# Terminología estándar del análisis de componentes principales

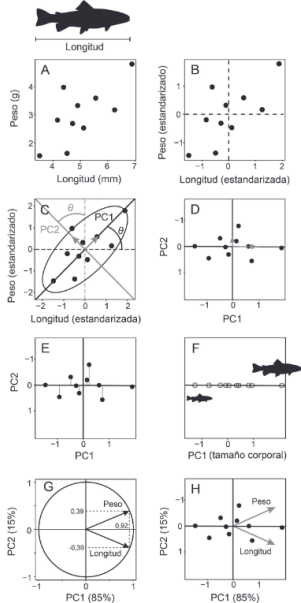
---

**Biplot:** gráfico en el cual se representan dos conjuntos de objetos de diferentes formas (individuos y variables). Esta superposición es ficticia, dado que las nubes de las UE y las variables no se encuentran en el mismo espacio.

**La calidad de la representación** indica qué tan bien está proyectada una unidad experimental (UE) o variable en el espacio de ordenación. Se mide con el coseno cuadrado del ángulo entre una UE o variable y un componente principal (PC), equivalente al coeficiente de correlación al cuadrado ( $\text{loading}^2$ ), y varía entre 0 y 1. La suma de los valores de calidad a lo largo de todos los PCs es igual a 1 para cada UE o variable.

**Contribución:** importancia relativa de una UE o variable en la construcción de un determinado PC. La suma de todas las contribuciones de las UE o variables para un mismo PC es igual a 100.





**Fig. 6.1. PCA.** (A) Gráfico de dispersión de 10 individuos de una especie de pez y dos variables (longitud y peso); (B) estandarización de las variables a media 0 y varianza 1; (C) elipse que engloba las UE y componentes principales (PC) 1 y 2, estos ejes corresponden a la rotación de la ordenada y la abscisa un cierto ángulo  $\theta$ , los vectores (flechas grises) representan los eigenvectores del PCA cuyas longitudes son 1; (D) resultado del PCA en dos dimensiones, se muestran los eigenvectores multiplicados por la raíz cuadrada de sus eigenvalores (*loadings*); (E) las UE se proyectan sobre el PC1 (líneas perpendiculares); (F) las nuevas coordenadas de las UE proyectadas sobre el PC1 (círculos blancos) se denominan *scores*; (G) círculo de correlación con radio igual a 1, donde se muestran las variables (vectores) y sus coordenadas (*loadings*) que definen la calidad de la representación; (H) *biplot* de UE vs. variables.

# ANÁLISIS MULTIVARIADO DEL GÉNERO BULNESIA

Variable	Estados	Codificación
1. Hábito	Arbustos	0
	Arbustos y árboles	1
	Árboles	2
2. Longitud del internodio (cm)	-	-
3. Diámetro del internodio (cm)	-	-
4. Longitud de la hoja (cm)	-	-
5. Ancho de la hoja (cm)	-	-
6. Longitud del peciolo (cm)	-	-
7. Número de folíolos	-	-
8. Presencia de peciólulos	Folíolos no sésiles	0
	Folíolos sésiles o no sésiles	1
	Folíolos sésiles	2
9. Disposición de los folíolos en el raquis	Folíolos alternos	0
	Folíolos subpuestos	1
	Folíolos opuestos	2
10. Pubescencia de la hoja	Ausente	0
	Ausente y presente	1
	Presente	2
11. Longitud del foliolo (mm)	-	-
12. Ancho del foliolo (mm)	-	-
13. Número de nervaduras primarias del foliolo	-	-

13. Número de nervaduras primarias del foliolo	-	-
14. Posición de los folíolos terminales	Paralelos	0
	Paralelos y divergentes	1
	Divergentes	2
15. Presencia de mucrón en folíolos	Folíolos no mucronados	0
	Folíolos mucronados	1
16. Tipo de inflorescencia	Flores solitarias	1
	Inflorescencia en dicasio	2
17. Longitud del pedúnculo (mm)	-	-
18. Longitud del sépalo (mm)	-	-
19. Ancho del sépalo (mm)	-	-
20. Color de los pétalos	Blanco	1
	Amarillo	2
21. Longitud del pétalo (mm)	-	-
22. Ancho del pétalo (mm)	-	-
23. Número de nervaduras del pétalo	-	-
24. Tipo de estambres	Heterogéneos	0
	Heterogéneos y homogéneos	1
	Homogéneos	2
25. Modificación de los estambres	No modificados	0
	Modificados y no modificados	1
	Modificados	2

Crisci et al. 1979

26. Presencia de gran escama junto al estambre	Ausente	0
	Ausente y presente	1
	Presente	2
27. Presencia de pelos en la base del filamento estaminal	Sin pelos	0
	Con o sin pelos	1
	Con pelos	2
28. Presencia de una escama suplementaria junto al estambre	Ausente	0
	Presente	1
29. Agrupación de los estambres	No agrupados	0
	Agrupados o no agrupados	1
	Agrupados	2
30. Longitud del filamento (mm)	-	-
31. Longitud de la antera (mm)	-	-
32. Longitud de la escama (mm)	-	-
33. Presencia de ápice laciniado en la escama estaminal	Sin ápice laciniado	0
	Con o sin ápice laciniado	1
	Con ápice laciniado	2
34. Número de carpelos	En número de 3	0
	En número de 3 y 5	1
	En número de 5	2
35. Curvatura del estilo	Estilo no curvado	0
	Estilo curvado o no curvado	1
	Estilo curvado	2

36. Número de óvulos por carpelo	-	-
37. Pubescencia del fruto	Glabro	0
	Pubescente	1
38. Longitud del fruto (mm)	-	-
39. Ancho del fruto (mm)	-	-
40. Desarrollo del carpóforo	Reducido	1
	Bien desarrollado	2
41. Longitud del carpóforo (mm)	-	-
42. Forma de la semilla	Semicircular o semielíptica	1
	Oblongo-reniforme	2
43. Longitud de la semilla (mm)	-	-

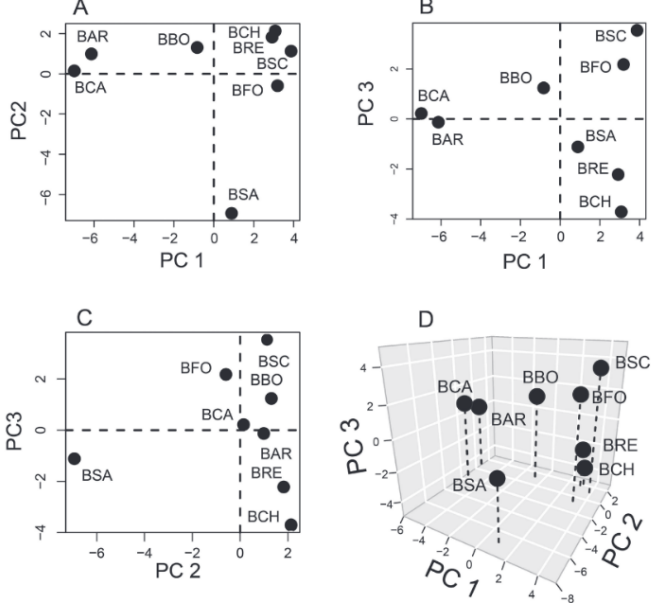
		Variables				
		1	2	3	...	$p$
UE	1	$x_{11}$	$x_{12}$	$x_{13}$	...	$x_{1p}$
	2	$x_{21}$	$x_{22}$	$x_{23}$	...	$x_{2p}$
	3	$x_{31}$	$x_{32}$	$x_{33}$	...	$x_{3p}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	$n$	$x_{n1}$	$x_{n2}$	$x_{n3}$	...	$x_{np}$

Especie	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21
<i>B. arborea</i>	2	35	2,1	85	57	7,7	13	2	0	2	30	8,6	6	0	1	2	17	7,1	3,4	2	22
<i>B. carrapo</i>	2	36	1,6	97	71	9	7	2	0	2	40	16	6	1	1	2	18	6,4	5,8	2	24
<i>B. chilensis</i>	0	24	2,6	14	8,9	1,8	8	2	1	0	5,2	2,4	NA	2	1	1	9	7,1	4,2	2	9,8
<i>B. bonariensis</i>	0	20	1,3	26	18	3,4	14	1	0	2	8,9	2	1	2	1	1	12	6,8	3,9	2	18
<i>B. retama</i>	1	40	2	13	11	3,1	5	1	1	2	6,6	2,6	2	1	2	1	10	7,4	4,4	2	7,7
<i>B. foliosa</i>	0	19	1,3	28	25	5,8	4	1	1	2	14	7,8	3	2	1	1	13	5,3	3	2	8,5
<i>B. schickendantzii</i>	0	10	1,9	20	12	2,7	10	0	0	2	5,7	1,9	1	2	1	1	10	5,9	3,1	2	9,3
<i>B. sarmientoi</i>	2	22	1,4	21	27	5,1	2	2	2	1	17	12	5	2	1	1	3,9	2,9	2,3	1	12

Especie	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C38	C39	C40	C41	C42	C43
<i>B. arborea</i>	17	16	0	1	2	0	1	2	10	1,5	5,2	2	2	1	2	1	46	41	2	7,7	1	13
<i>B. carrapo</i>	19	12	0	2	2	0	1	2	9,9	1,4	53	2	2	2	2	1	56	52	2	5,3	1	12
<i>B. chilensis</i>	5,7	8	0	0	0	2	1	2	7	2	4,4	1	1	NA	7	1	13	12	2	0,6	2	2,7
<i>B. bonariensis</i>	10	10	0	1	1	0	1	1	11	1,6	4,4	0	2	2	1	1	36	33	2	4,8	1	11
<i>B. retama</i>	4,6	7	1	0	0	1	1	1	7,1	2,2	3,1	1	2	1	8	1	23	19	1	0,8	2	11
<i>B. foliosa</i>	2,7	6	2	0	0	0	1	0	6,2	1,6	3,9	1	2	1	4	2	16	13	1	0,7	2	4,9
<i>B. schickendantzii</i>	4,3	5	1	0	0	0	2	0	7,4	1,7	4,4	2	2	1	4	2	12	13	1	0,4	2	5,3
<i>B. sarmientoi</i>	7	6	2	0	0	0	1	0	4,1	1,1	2,9	2	0	0	2	1	52	48	2	5,2	1	14

**Tabla 6.1.** Tabla resumen del PCA para la MBD de especies de *Bulnesia*. Se muestran los primeros siete componentes principales, sus eigenvalores ( $\lambda_i$ ), el porcentaje de variación explicada por cada componente y el porcentaje acumulado.

<b>Componente principal</b>	<b>Eigenvalor</b>	<b>Porcentaje de variación explicada</b>	<b>Porcentaje acumulado</b>
PC1	18,731	45,685	45,685
PC2	8,620	21,024	66,719
PC3	5,556	13,536	80,246
PC4	3,353	8,189	88,414
PC5	2,393	5,847	94,251
PC6	1,455	3,535	97,786
PC7	0,912	2,224	99,999

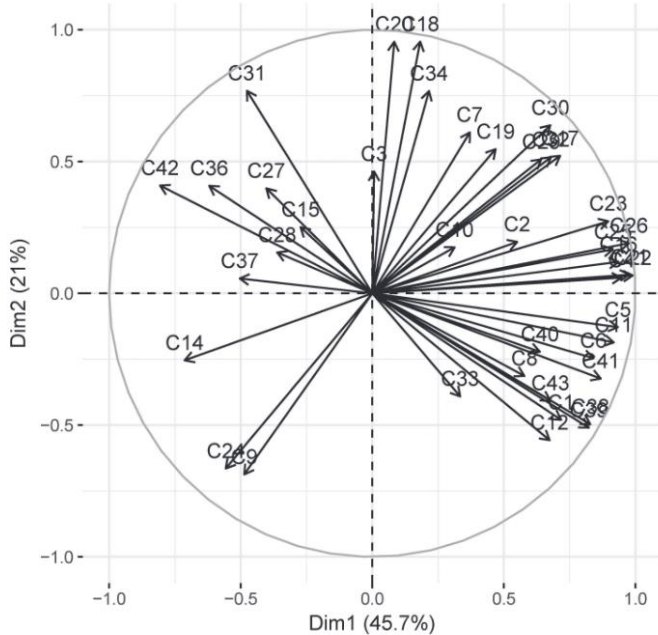


**Fig. 6.2.** Visualización del PCA aplicado a la MBD de especies de *Bulnesia*. Se grafican los espacios de ordenación bidimensionales del (A) PC1 vs. PC2; (B) PC1 vs. PC3; (C) PC2 vs. PC3; (D) espacio de ordenación tridimensional para el PC1, PC2 y PC3. BAR: *B. arborea*, BCA: *B. carrapo*, BCH: *B. chilensis*, BBO: *B. bonariensis*, BRE: *B. retama*, BFO: *B. foliosa*, BSC: *B. schickendantzii*, BSA: *B. sarmiento*.

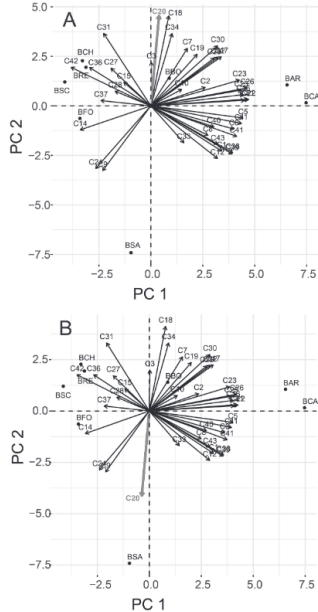
**Tabla 6.2. Loadings de los tres primeros componentes principales (PCs) para cada una de las variables de la MBD de especies de *Bulnesia*. Las variables 13 y 35 fueron excluidas del análisis por presentar valores no disponibles.**

Variable	PC1	PC2	PC3
1. Hábito	0,717	-0,48	0,262
2. Longitud del internodio	0,551	0,195	0,562
3. Diámetro del internodio	0,007	0,462	0,759
4. Longitud de la hoja	0,947	0,059	-0,132
5. Ancho de la hoja	0,933	-0,128	-0,099
6. Longitud del peciolo	0,840	-0,246	-0,193
7. Número de folíolos	0,371	0,611	-0,261
8. Presencia de peciúlos	0,578	-0,314	0,675
9. Disposición de los folíolos en el raquis	-0,487	-0,687	0,488
10. Pubescencia de la hoja	0,313	0,175	-0,696
11. Longitud del foliolo	0,917	-0,186	-0,062
12. Ancho del foliolo	0,674	-0,557	-0,007
14. Posición de los foliolos terminales	-0,713	-0,254	-0,181
15. Presencia de mucrón en foliolos	-0,273	0,251	0,381
16. Tipo de inflorescencia	0,936	0,119	-0,012
17. Longitud del pedúnculo	0,713	0,522	-0,287
18. Longitud del sépalo	0,182	0,955	0,192
19. Ancho del sépalo	0,468	0,547	0,269
20. Color de los pétalos	0,083	0,955	-0,191
21. Longitud del pétalo	0,977	0,071	-0,088
22. Ancho del pétalo	0,990	0,068	0,036
23. Número de nervaduras del pétalo	0,897	0,273	0,133
24. Tipo de estambres	-0,557	-0,665	-0,236
25. Modificación de los estambres	0,914	0,167	-0,124
26. Presencia de gran escama junto al estambre	0,975	0,19	-0,094
27. Presencia de pelos en la base del filamento estaminal	-0,403	0,398	0,786
28. Presencia de una escama suplementaria junto al estambre	-0,361	0,155	-0,607
29. Agrupación de los estambres	0,642	0,508	0,539
30. Longitud del filamento	0,676	0,638	-0,198
31. Longitud de la antera	-0,476	0,768	0,343
32. Longitud de la escama	0,679	0,516	-0,248
33. Presencia de ápice laciniado en la escama estaminal	0,334	-0,391	-0,104
34. Número de carpelos	0,216	0,768	-0,485
36. Número de óvulos por carpelo	-0,620	0,408	0,516
37. Pubescencia del fruto	-0,504	0,056	-0,749
38. Longitud del fruto	0,828	-0,497	0,094
39. Ancho del fruto	0,823	-0,51	0,072
40. Desarrollo del carpóforo	0,637	-0,222	0,41
41. Longitud del carpóforo	0,868	-0,324	0,016
42. Forma de la semilla	-0,807	0,409	0,024
43. Longitud de la semilla	0,677	-0,408	0,083





**Fig. 6.3.** Círculo de correlación unitario aplicado a la MBD de *Bulnesia* que muestra la calidad de la representación de las variables. Los códigos corresponden a los caracteres de la Tabla 2.8.



**Fig. 6.4.** (A) *Biplot* aplicado a la MBD de *Bulnesia*, observe que la ubicación de las UE en el espacio es una imagen especular de la expuesta en la Figura 6.2, (B) *biplot* resultado de la recodificación de la variable categórica C20 (color de los pétalos). Los códigos de las variables corresponden a los caracteres de la Tabla 2.8, BAR: *B. arborea*, BCA: *B. carrapo*, BCH: *B. chilensis*, BBO: *B. bonariensis*, BRE: *B. retama*, BFO: *B. foliosa*, BSC: *B. schickendantzii*, BSA: *B. sarmentoi*.

---

Gracias

