

# Análisis de Componentes Principales (PCA)



José A. Perusquía Cortés  
Análisis Multivariado, Semestre 2025-II



# Introducción

- ▶ Motivación

Visualizar y/o interpretar datos multivariados es complicado

- ▶ A grandes rasgos PCA es un método estadístico que busca

1. “Reducir” la dimensionalidad de los datos

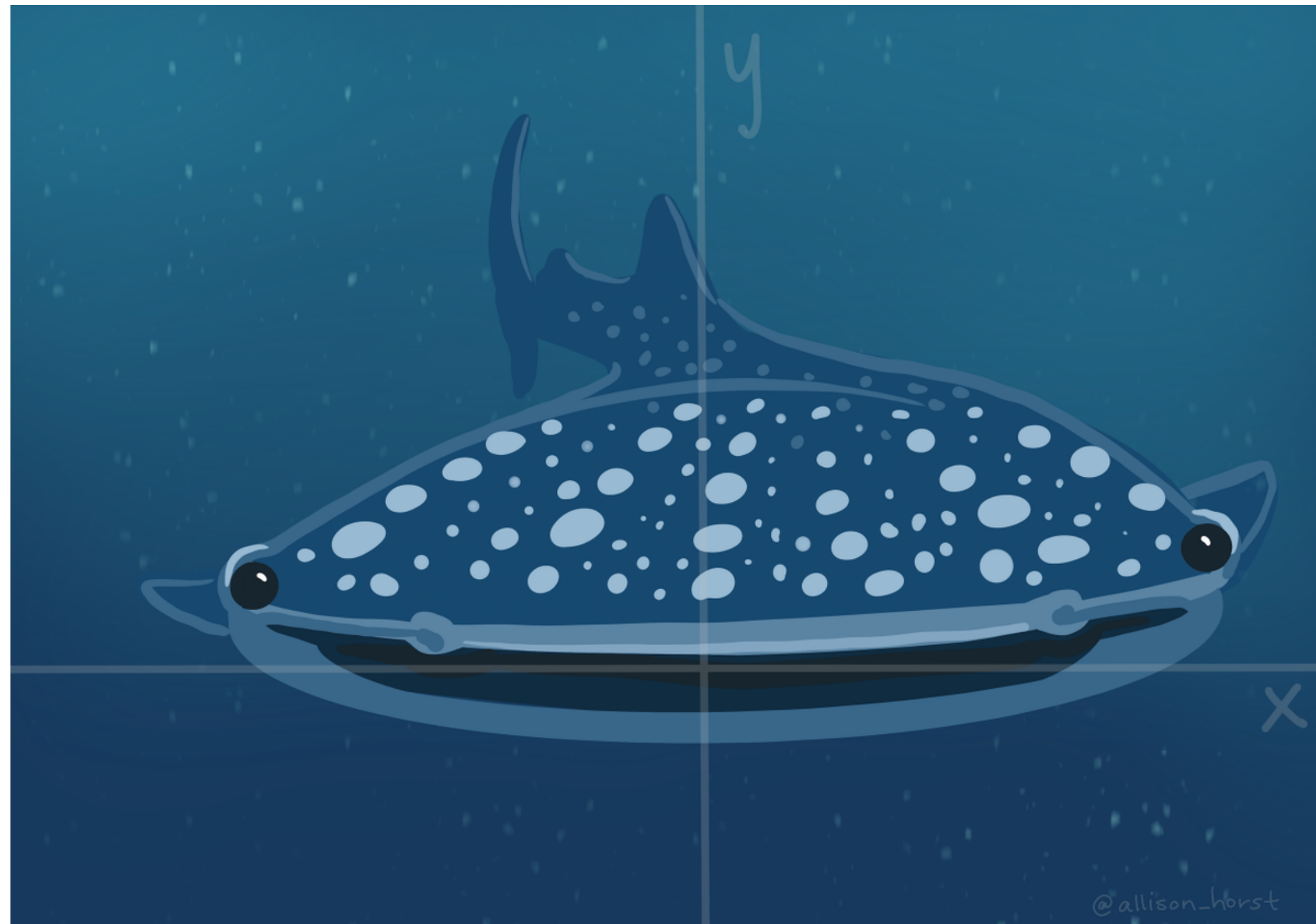
2. Retener la mayor cantidad de la variación original

- ▶ ¿Cómo?

Crear un nuevo conjunto de variables no correlacionadas y ordenadas por varianza

# Intuición

¿Cómo debe girar la cabeza la ballena para comer la mayo cantidad de kril?



# Construcción (idea)

- ▶ Sea  $\mathbf{x}_{p \times 1}$  un vector aleatorio real valuado

- ▶ El primer componente principal estará dado por:

$$\alpha_1^T \mathbf{x} = \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1p}x_p = \sum_{j=1}^p \alpha_{1j}x_j$$

tal que sea la combinación lineal de **mayor varianza**.

- ▶ El segundo componente principal estará dado por:

$$\alpha_2^T \mathbf{x} = \alpha_{21}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{2p}x_p = \sum_{j=1}^p \alpha_{2j}x_j$$

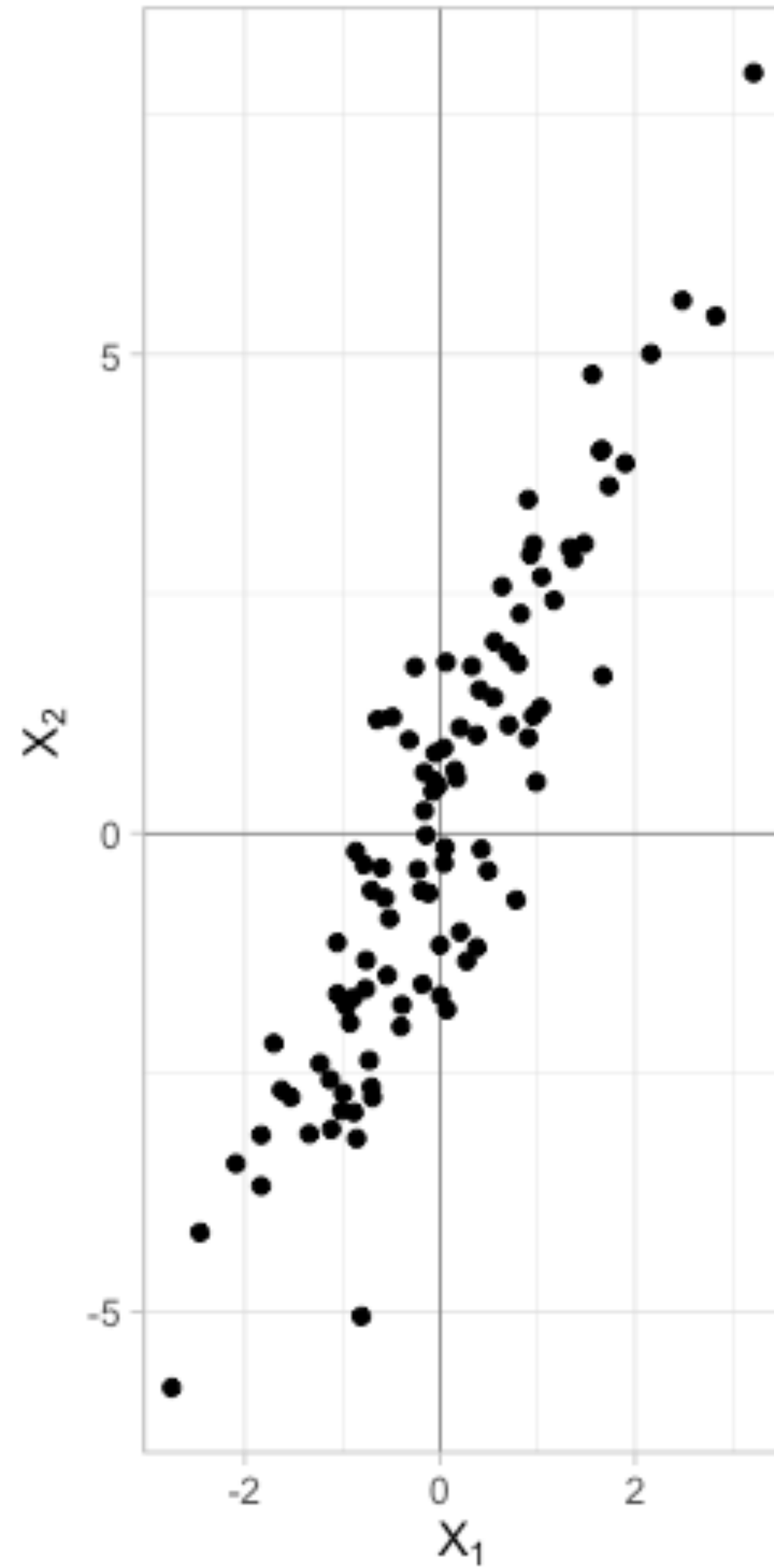
tal que sea la combinación lineal de **mayor varianza** y **no esté correlacionado** con el primero.

- ▶ Y así sucesivamente...

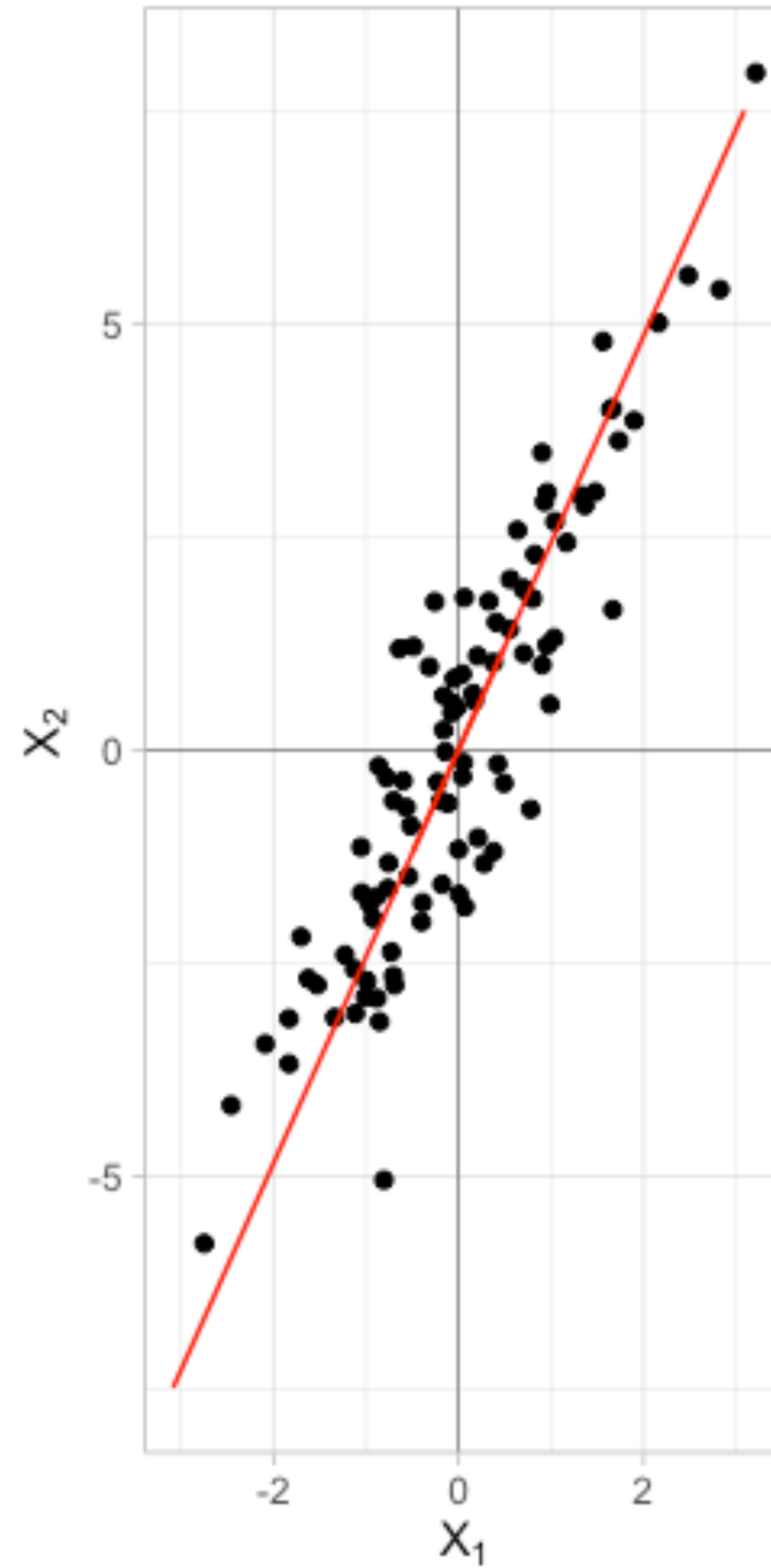


# Ejemplo

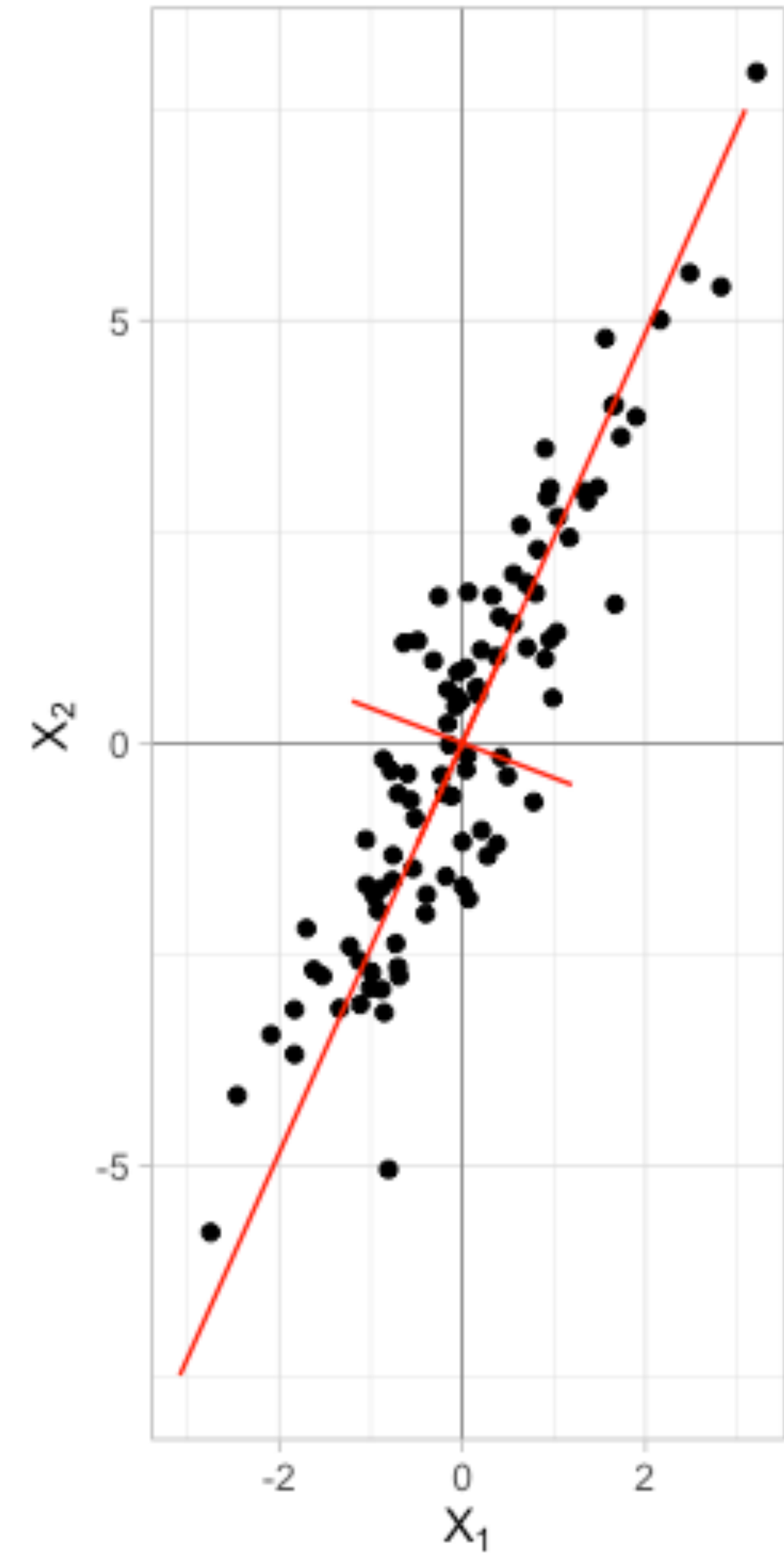
Datos



Primer Componente

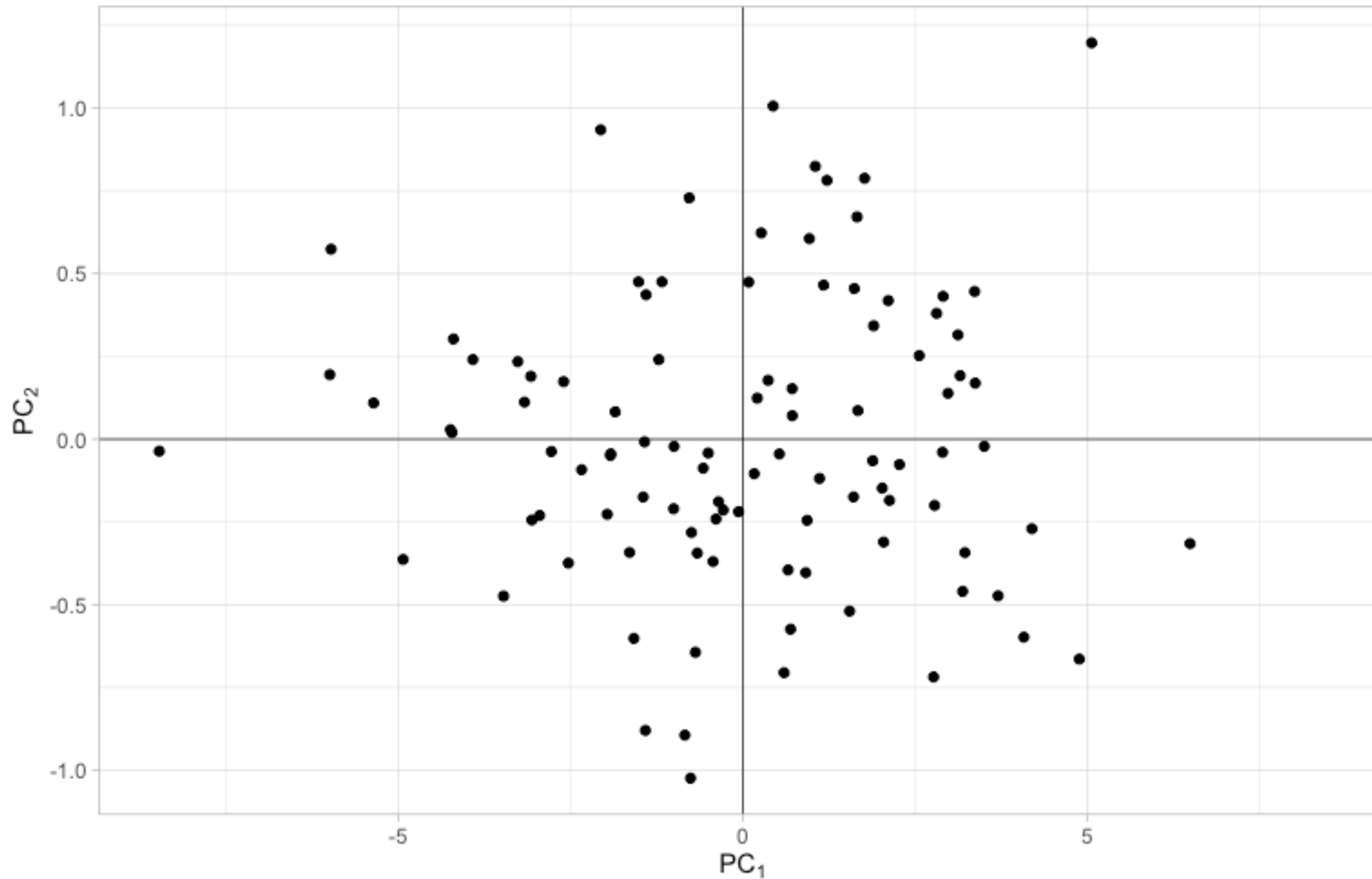


Segundo Componente



# Ejemplo

- Nuevas variables



# Construcción

- ▶ Sea  $\mathbf{x}_{p \times 1}$  un vector aleatorio real valuado, con  $\Sigma$  conocida
- ▶ (Formalmente) el primer componente principal se encuentra resolviendo

$$\begin{aligned} \max_{\alpha_1} \quad & \text{var}(\alpha_1^T \mathbf{x}) = \alpha_1^T \Sigma \alpha_1 \\ \text{s.a.} \quad & \alpha_1^T \alpha_1 = 1 \end{aligned}$$

- ▶ Se resuelve con los multiplicadores de Lagrange
- ▶ Dando como resultado que
  - $\lambda$ : eigenvalor más grande
  - $\alpha_1$ : eigenvector asociado

# Construcción

- El segundo componente se encuentra resolviendo:

$$\begin{aligned} \max_{\alpha_2} \quad & \text{var}(\alpha_2^T x) = \alpha_2^T \Sigma \alpha_2 \\ \text{s.a.} \quad & \alpha_2^T \alpha_2 = 1 \\ & \text{cov}(\alpha_1^T \mathbf{X}, \alpha_2^T \mathbf{X}) = 0 \end{aligned}$$

- Se resuelve con los multiplicadores de Lagrange
- Dando como resultado que
  - $\lambda$  : segundo eigenvalor más grande
  - $\alpha_2$  : eigenvector asociado
- Y así sucesivamente...



# Interpretación geométrica

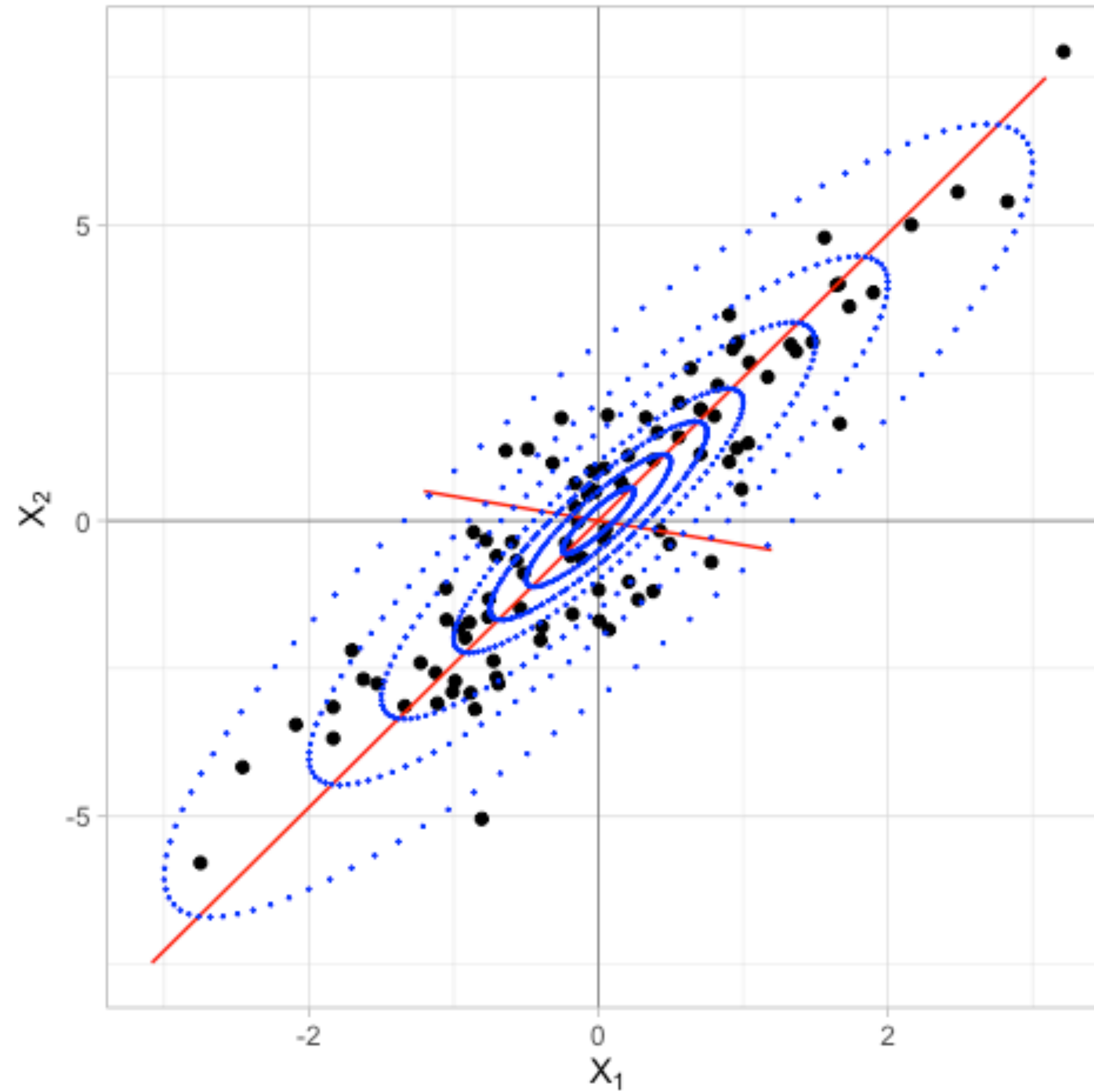
## Observación 1

Los componentes principales corresponden a una transformación ortogonal de  $\mathbf{x}$ , dada por  $\mathbf{z} = \mathbf{A}^T \mathbf{x}$ , donde  $\mathbf{A}$  es la matriz de vectores propios y así  $\text{var}(z_k) = \lambda_k$ .

## Proposición 1

Sea la familia de elipsoides dada por  $\mathbf{x}^T \Sigma^{-1} \mathbf{x} = c$ , entonces los componentes principales definen los ejes principales.

# Interpretación geométrica



# Propiedades algebraicas

## Propiedad 1

Sea la transformación ortogonal  $\mathbf{y} = \mathbf{B}^T \mathbf{x}$ , donde  $\mathbf{B}_{q \times p}$  y  $\Sigma_y = \mathbf{B}^T \Sigma \mathbf{B}$  entonces,

1.  $\text{tr}(\Sigma_y)$  y  $|\Sigma_y|$  se maximizan cuando  $\mathbf{B} = \mathbf{A}_q$  (las primeras  $q$  columnas)
2.  $\text{tr}(\Sigma_y)$  se minimiza cuando  $\mathbf{B} = \mathbf{A}_q^*$  (las últimas  $q$  columnas)

## Propiedad 2

La descomposición espectral de  $\Sigma$  está dada por  $\Sigma = \sum_{i=1}^p \lambda_i \alpha_i \alpha_i^T$ .

## Propiedad 3

Si  $\sigma_j^2$  es la varianza residual de predecir  $x_j$  en términos de  $\mathbf{y}$ , entonces  $\sum \sigma_i^2$  se minimiza cuando  $\mathbf{B} = \mathbf{A}_q$ .

# Componentes principales muestrales

- Sea  $\mathbf{x}_1, \dots, \mathbf{x}_n$  una muestra aleatoria (centrada) con matriz de covarianzas muestral

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X}$$

entonces:

1. El primer componente es el vector propio  $\mathbf{a}_1$  asociado al valor propio más grande de  $\mathbf{S}$
2. Se tienen  $n$  nuevas variables  $z_{i1} = \mathbf{a}_1^T \mathbf{x}_i$
3. Y sucesivamente para los otros componentes

## Observación 2

1. A los vectores  $\{\mathbf{z}_i\}_i$  se les conoce como **scores**
2. A los vectores propios  $\{\mathbf{a}_i\}_i$  se les conoce como **loadings**

# Consideraciones prácticas

## Observación 3

Computacionalmente es preferible utilizar la descomposición en valores singulares (SVD) para encontrar los componentes principales

$$\mathbf{S} = \frac{1}{n-1} \mathbf{W}^T \mathbf{W}$$

$$\mathbf{W} = \mathbf{U} \mathbf{D} \mathbf{V}^T$$

$$\mathbf{S} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T.$$

Esto debido a que:

1. Numéricamente más estable.
2. Permite considerar el caso  $p > n$ .
3. Puede ser más rápido.



# Consideraciones prácticas

## Proposición 2

Sea  $\mathbf{Z} = \mathbf{H}\mathbf{X}\mathbf{V}$  la matriz de cargas, i.e.,  $\mathbf{z}_i = \mathbf{V}^T(\mathbf{x}_i - \bar{\mathbf{x}})$  entonces se cumple:

1. La media muestral es el vector de ceros
2. La matriz de covarianzas es  $\Lambda$
3.  $\mathbf{v}_1^T \mathbf{S} \mathbf{v}_1 > \mathbf{v}_2^T \mathbf{S} \mathbf{v}_2 > \dots > \mathbf{v}_p^T \mathbf{S} \mathbf{v}_p$  y si  $\text{ran}(S) = q < p$  se tiene que  $\mathbf{v}_s^T \mathbf{S} \mathbf{v}_s = 0$  para

$$s = q + 1, \dots, p$$

$$4. \sum_{i=1}^p \mathbf{v}_i^T \mathbf{S} \mathbf{v}_i = \sum_{i=1}^p \lambda_i = \text{tr}(S)$$

$$5. \prod_{i=1}^p \mathbf{v}_i^T \mathbf{S} \mathbf{v}_i = \prod_{i=1}^p \lambda_i = |S|$$

# Consideraciones prácticas

## **Observación 4**

Dependiendo de la naturaleza de los datos muchas veces es preferible definir a los componentes principales con variables estandarizadas. Esto permite que:

1. Todas las propiedades sigan siendo válidas
2. Se pueden mezclar variables en diferentes escalas
3. Los componentes no estén dominados por una posible variable de mayor varianza

# Ejemplo: Calificaciones

- ▶ 88 calificaciones de 5 exámenes a libro abierto o cerrado

Alumno	Lineal (C)	Estadística (C)	Probabilidad(A)	Finanzas (A)	Cálculo (A)
1	97	92	77	72	96
2	83	88	90	75	96
3	95	83	81	71	96
4	75	82	73	75	83
5	83	73	75	75	78
.	.	.	.	.	.
.	.	.	.	.	.
87	25	36	25	25	35
88	20	50	31	14	29

- ▶ En R usamos `prcomp()` con  $\hat{\Sigma} = S$

# Ejemplo: Calificaciones

- Los valores propios (varianzas de los componentes) resultantes son:

$$\lambda_1 = 689.6583 > \lambda_2 = 200.9016 > \lambda_3 = 103.5280 > \lambda_4 = 83.3404 > \lambda_5 = 32.2476$$

- Los vectores de cargas:

<b>Lineal (C)</b>	-0.502	-0.759	0.289	-0.284	-0.080
<b>Estadística (C)</b>	-0.371	-0.188	-0.417	0.785	-0.186
<b>Probabilidad (A)</b>	-0.345	0.077	-0.144	-0.002	0.923
<b>Finanzas (A)</b>	-0.450	0.299	-0.591	-0.523	-0.287
<b>Cálculo (A)</b>	-0.535	0.541	0.609	0.164	-0.149

# Ejemplo: Calificaciones

- ▶ El primer componente es un “promedio”

$$-0.502 \cdot \text{Lineal} - 0.371 \cdot \text{Estadística} - 0.345 \cdot \text{Proba.} - 0.450 \cdot \text{Finanzas} - 0.535 \cdot \text{Cálculo}$$

- ▶ El segundo componente es una comparación entre libro abierto y cerrado

$$-0.759 \cdot \text{Lineal} - 0.188 \cdot \text{Estadística} + 0.077 \cdot \text{Proba.} + 0.299 \cdot \text{Finanzas} + 0.541 \cdot \text{Cálculo}$$

- ▶ El tercer componente es una comparación entre matemáticas “puras y aplicadas”

$$0.289 \cdot \text{Lineal} - 0.417 \cdot \text{Estadística} - 0.144 \cdot \text{Proba.} - 0.591 \cdot \text{Finanzas} + 0.609 \cdot \text{Cálculo}$$

- ▶ El cuarto componente (¿?)

$$-0.284 \cdot \text{Lineal} + 0.785 \cdot \text{Estadística} - 0.002 \cdot \text{Proba.} - 0.523 \cdot \text{Finanzas} + 0.164 \cdot \text{Cálculo}$$

- ▶ El quinto componente (¿?)

$$-0.080 \cdot \text{Lineal} - 0.186 \cdot \text{Estadística} + 0.923 \cdot \text{Proba.} - 0.287 \cdot \text{Finanzas} - 0.149 \cdot \text{Cálculo}$$



# Otras consideraciones

- La interpretación de los componentes requiere conocimiento del problema
  - Algunos componentes pueden interpretarse como un promedio ponderado
  - Algunos componentes pueden discriminar entre grupos de variables
  - Esto no siempre se cumple, a veces no es posible encontrarles significado
- **¿Cuántos componentes elegir?**

# Número de componentes

- ▶ Seleccionar los componentes que expliquen un cierto porcentaje de la variación (por ejemplo, 70%, 80%, 90%, etc.)
- ▶ Usar la regla de codo
- ▶ Otros (e.g. pruebas de hipótesis)
- ▶ **Regla de Kaiser.** Retener los componentes con varianza mayor a cierto valor (e.g.  $>.7$ )  
(para PCA con matriz de correlación)

# Ejemplo: Calificaciones

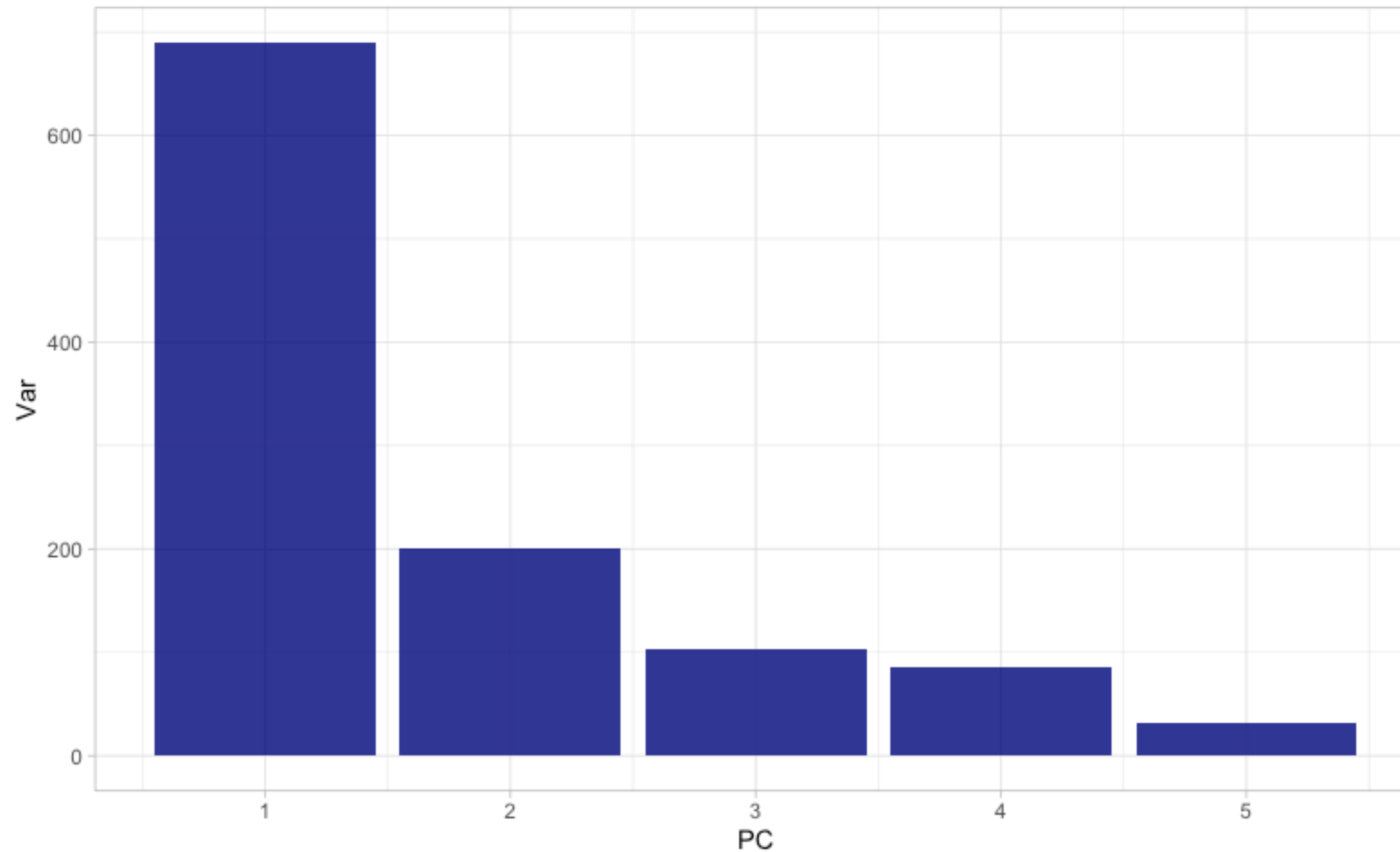
- La variación explicada por los componentes

61.91%   18.21%   9.35%   7.63%   2.90%

- Nos quedamos con los primeros dos para tener arriba del 80% de la variación total (80.12%)
- Nos quedamos con los primeros tres para tener casi 90% de la variación total (89.47%)

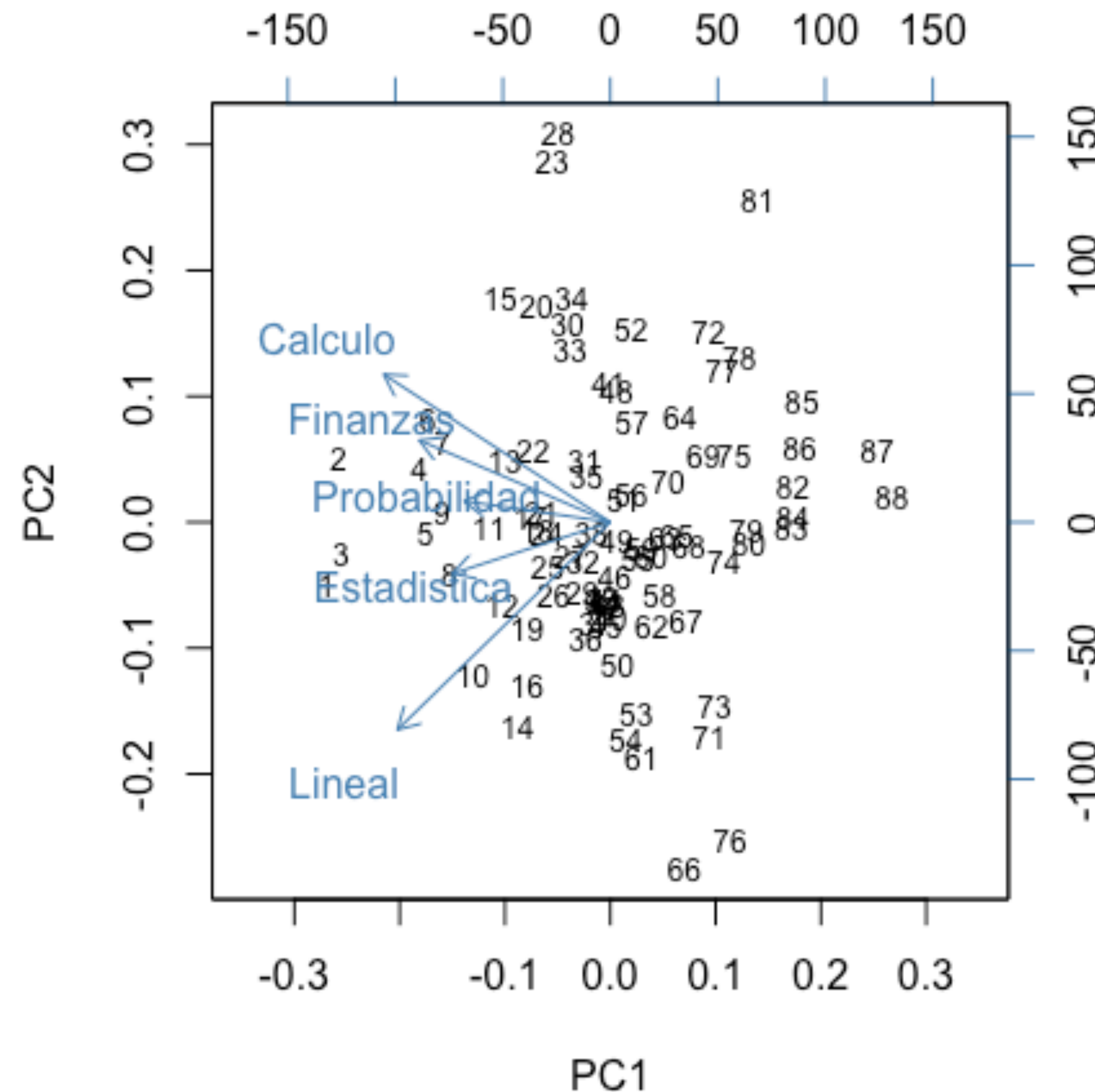
# Ejemplo: Calificaciones

- Regla de codo: graficar las varianzas (en **R** función **screeplot**)



# Ejemplo: Calificaciones

- Si nos quedamos con dos componentes podemos graficarlos usando `biplot()`





# Ejemplo: Calificaciones

- Los valores propios resultantes de hacer PCA con la matriz de correlación

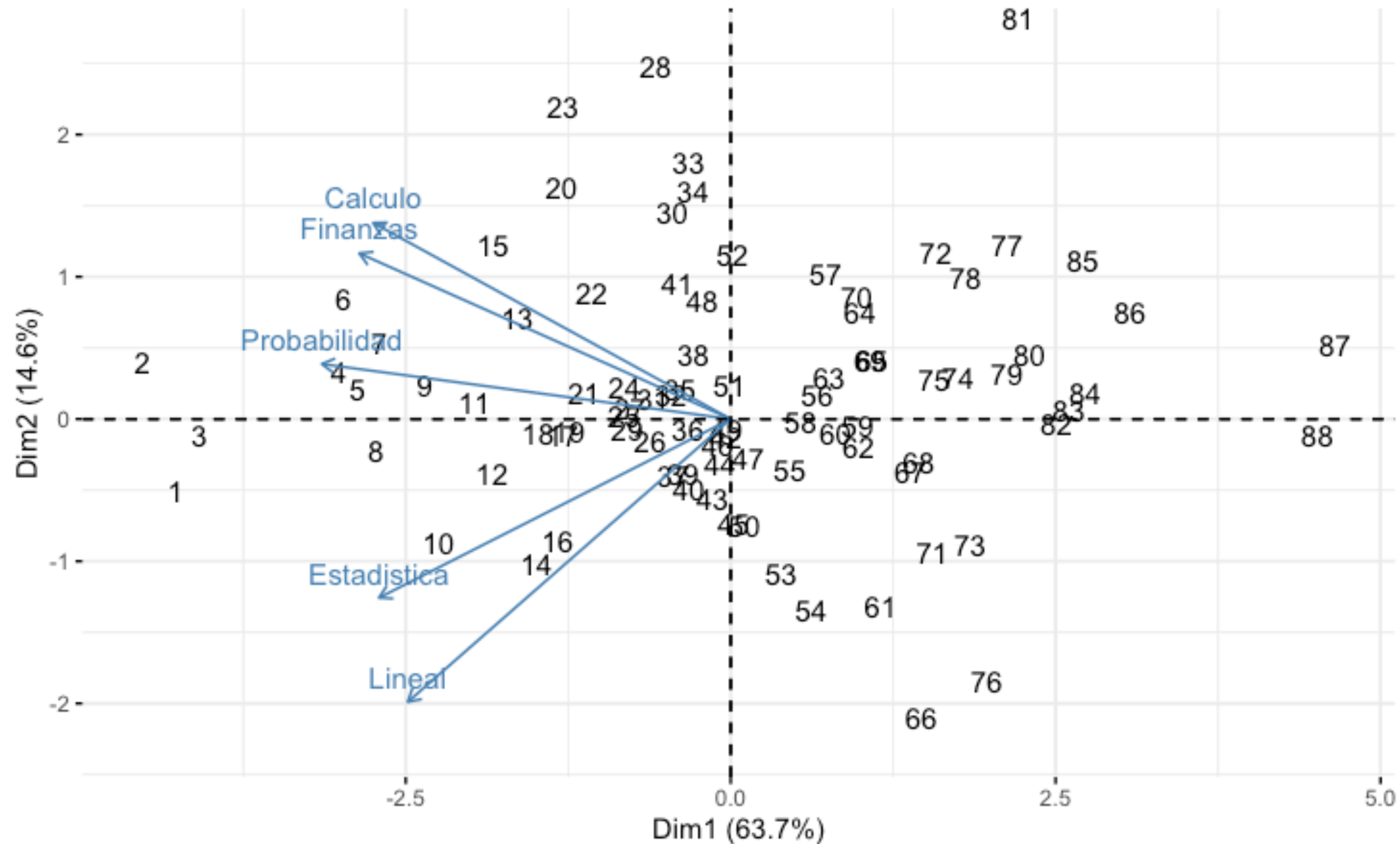
$$\lambda_1 = 3.1862 > \lambda_2 = 0.7287 > \lambda_3 = 0.4473 > \lambda_4 = 0.3917 > \lambda_5 = 0.2461$$

- Los vectores de cargas:

<b>Lineal (C)</b>	-0.397	-0.664	0.612	-0.091	-0.131
<b>Estadística (C)</b>	-0.432	-0.420	-0.740	0.234	-0.179
<b>Probabilidad (A)</b>	-0.502	0.129	-0.021	-0.116	0.846
<b>Finanzas (A)</b>	-0.456	0.389	-0.064	-0.674	-0.425
<b>Cálculo (A)</b>	-0.439	0.461	0.268	0.684	-0.230

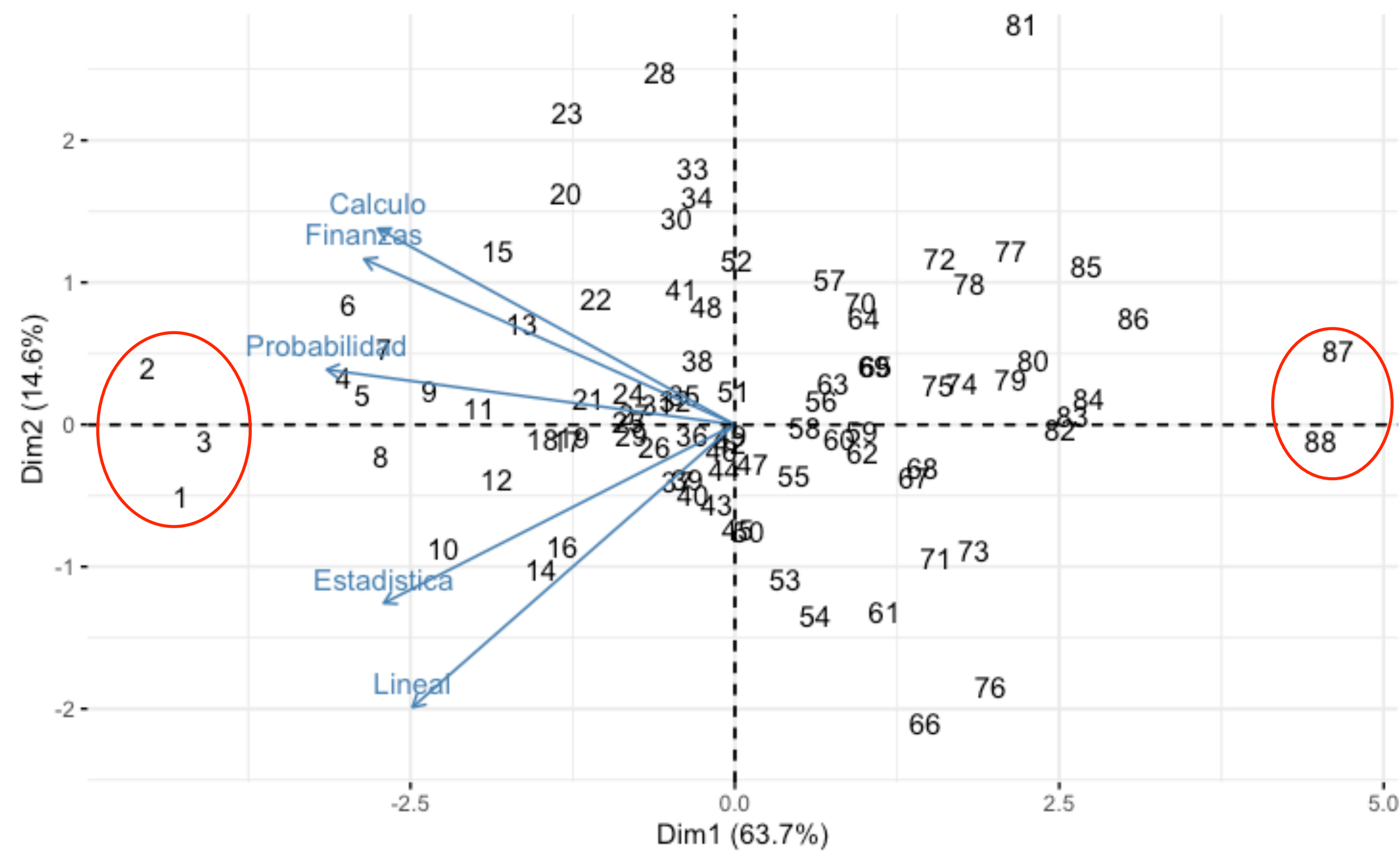
# Ejemplo: Calificaciones

- La librería **factoextra** proporciona una alternativa utilizando **ggplot**



# Ejemplo: Calificaciones

- Con el PC1, se pueden identificar los mejores y peores promedios

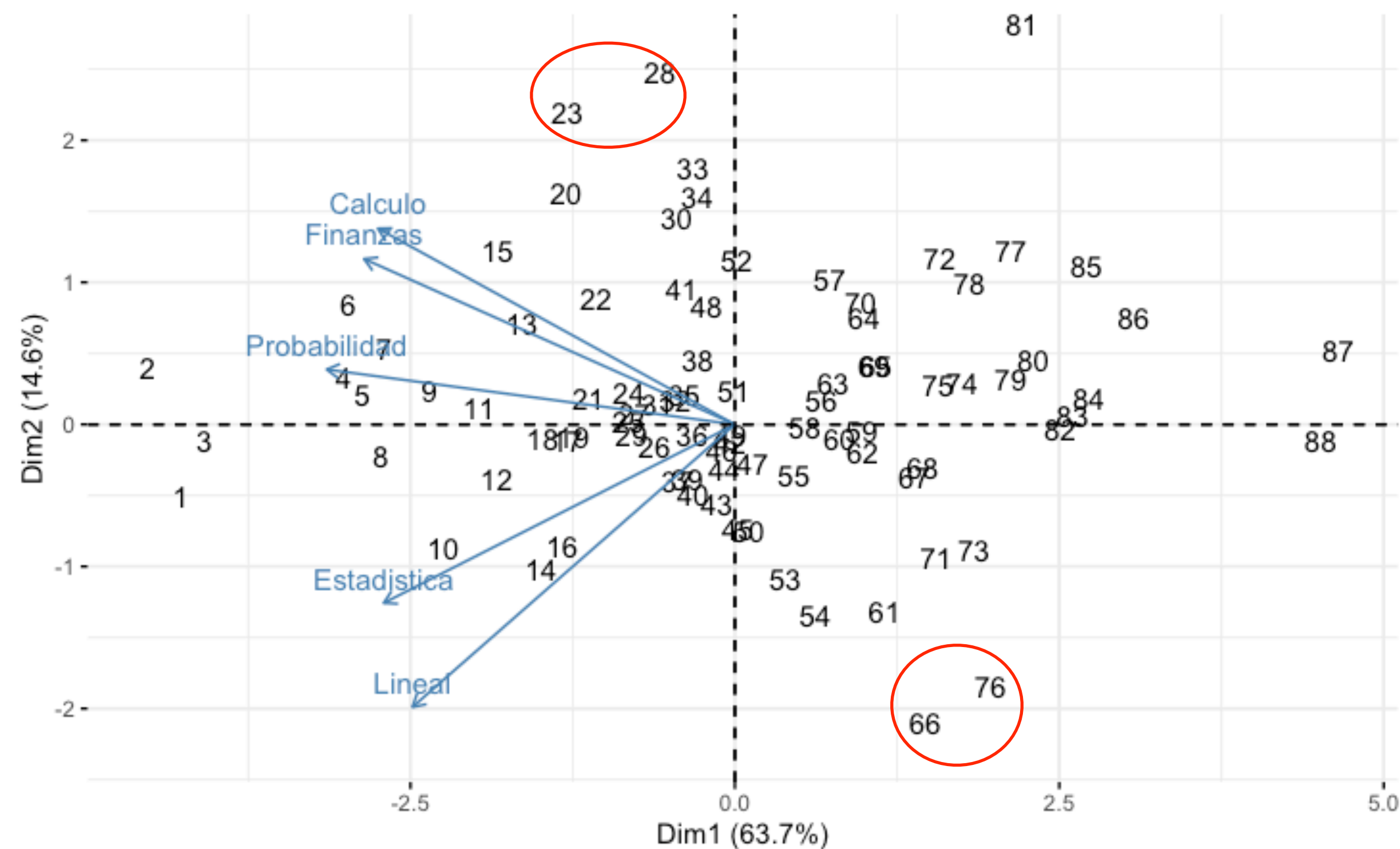


Alumno	Lineal	Est.	Proba.	Finanzas	Cálculo
1	97	92	77	72	96
2	83	88	90	75	96
3	95	83	81	71	96

Alumno	Lineal	Est.	Proba.	Finanzas	Cálculo
87	25	36	25	25	35
88	20	50	31	14	29

# Ejemplo: Calificaciones

- Con el PC2, se pueden identificar las mejores y peores calificaciones en examen abierto y cerrado



Alumno	Lineal	Est.	Proba.	Finanzas	Cálculo
66	79	63	47	27	34
76	69	60	48	28	24

Alumno	Lineal	Est.	Proba.	Finanzas	Cálculo
23	38	54	60	62	96
28	32	68	72	68	82

# Ejemplo: Pizzas

- ▶ 300 observaciones de 7 valores nutricionales en 10 marcas de pizza diferentes

1. **Mois:** Cantidad de agua por cada 100g
2. **Prot:** Cantidad de proteína por cada 100g
3. **Fat:** Cantidad de grasa por cada 100g
4. **Ash:** Cantidad de ceniza por cada 100g
5. **Sodium:** Cantidad de sodio por cada 100g
6. **Carb:** Cantidad de carbohidratos por cada 100g
7. **Cal:** Cantidad de calorías por cada 100g

- ▶ PCA con matriz de covarianzas, `prcomp(...,scale=F)`



# Ejemplo: Pizzas

- La raíz cuadrada de los valores propios (desviación estándar) resultantes

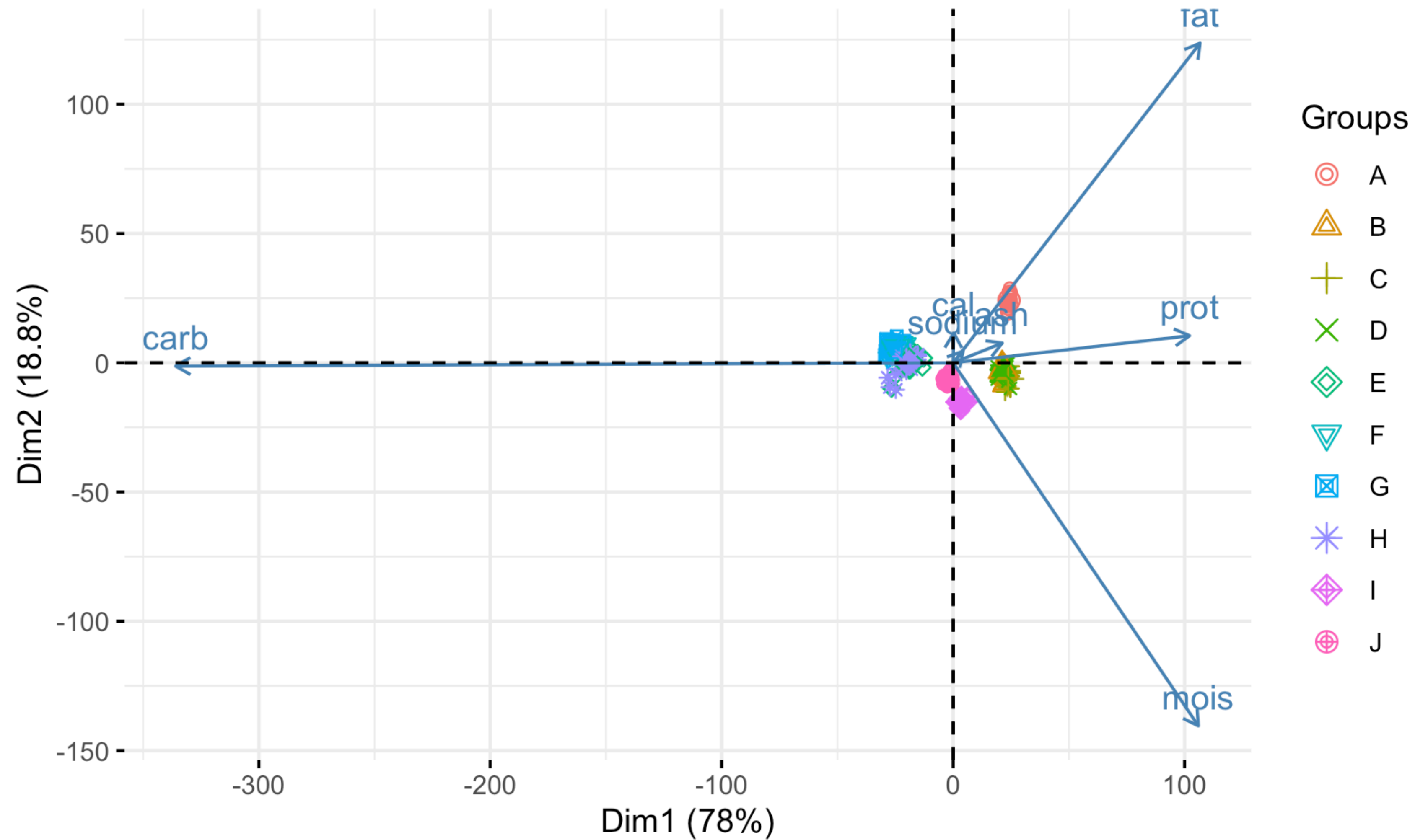
$$\lambda_1 = 20.53 > \lambda_2 = 10.09 > \lambda_3 = 4.10 > \lambda_4 = 0.42 > \lambda_5 = 0.09 > \lambda_6 = 0.03 > \lambda_7 = 0.01$$

- Los vectores de cargas:

<b>Moisture</b>	0.2769	-0.7470	0.3520	0.1958	0.0594	-0.4409	0.0698
<b>Protein</b>	0.2669	0.0557	-0.8097	0.2557	0.0837	-0.4434	0.0299
<b>Fat</b>	0.2789	0.6578	0.4679	0.2598	0.0357	-0.4486	-0.0198
<b>Ash</b>	0.0554	0.0406	-0.0222	-0.8714	-0.1666	-0.4502	0.0700
<b>Sodium</b>	0.0111	0.0238	0.0262	-0.2014	0.9783	0.0304	0.0025
<b>Carbs</b>	-0.8780	-0.0068	0.0124	0.1645	0.0574	-0.4444	0.0299
<b>Calories</b>	0.0006	0.0612	0.0100	0.0406	0.0014	0.0804	0.9939

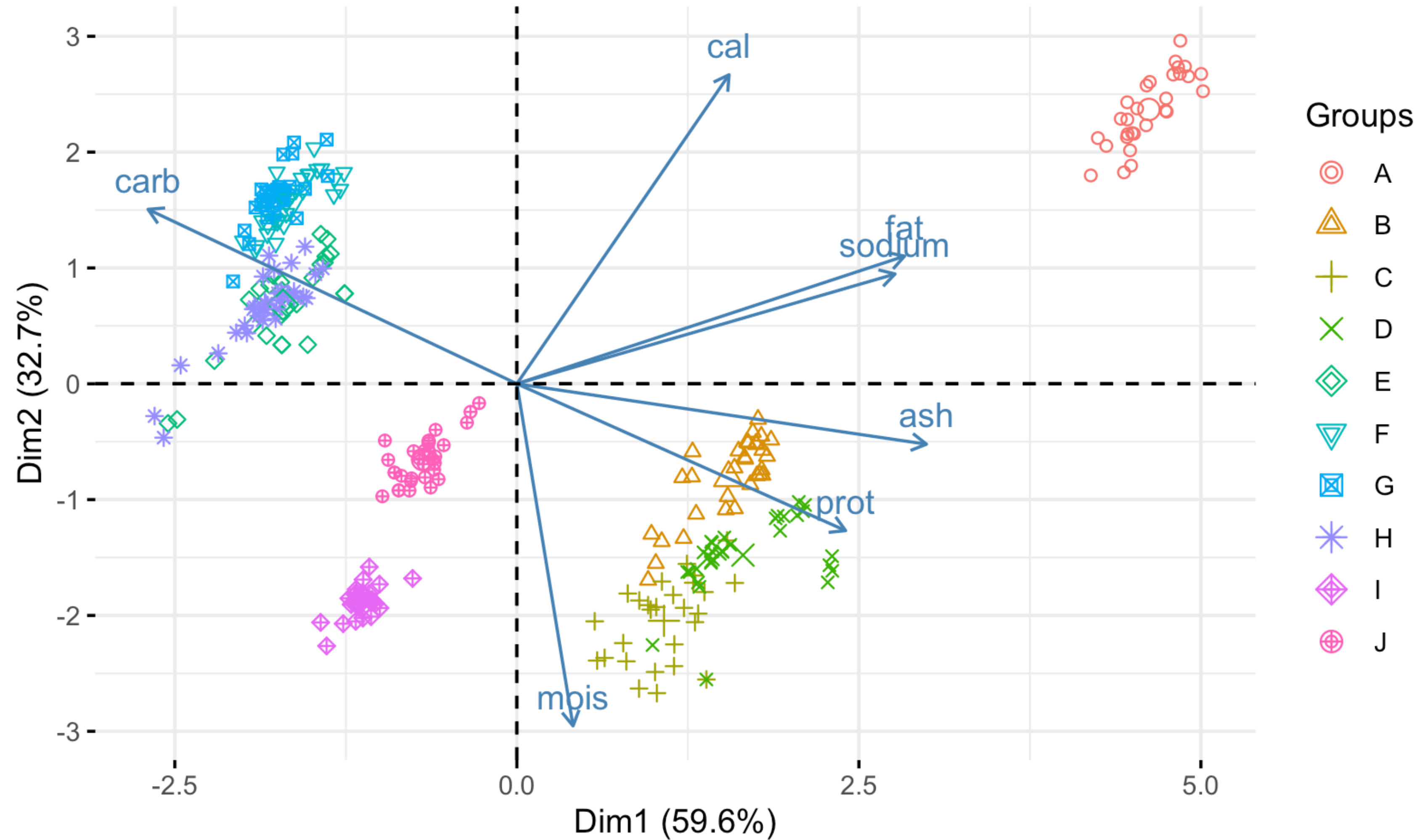
# Ejemplo: Pizzas

- Problemas por la escala/unidades/varianza de las variables



# Ejemplo: Pizzas

- PCA con matriz de correlación



# Ejemplo: Pizzas

- ▶ El primer componente es

$$0.064 \cdot \text{mois} + 0.378 \cdot \text{prot} + 0.446 \cdot \text{fat} + 0.471 \cdot \text{ash} \\ + 0.435 \cdot \text{sodium} - 0.424 \cdot \text{carb} + 0.244 \cdot \text{cal}$$

- ▶ El segundo componente es

$$-0.628 \cdot \text{mois} - 0.269 \cdot \text{prot} + 0.234 \cdot \text{fat} - 0.110 \cdot \text{ash} \\ + 0.201 \cdot \text{sodium} + 0.320 \cdot \text{carb} + 0.567 \cdot \text{cal}$$

- ▶ Ya no hay interpretación directa ni sencilla

# Aplicaciones y temas relacionados

- Selección de variables (problema NP-difícil)
- PCA + otros modelos/técnicas multivariadas (e.g. SVM, análisis de discriminantes, regresión, etc.)
- Detección de outliers y observaciones influyentes (analizando los primeros y los últimos componentes)
- Rotación de componente principales (para una mejor interpretación como en análisis de factores)
- Otro tipo de datos (e.g. series de tiempo, datos no independientes, discretos, etc.)