



UNIVERSIDAD DE
CÓRDOBA

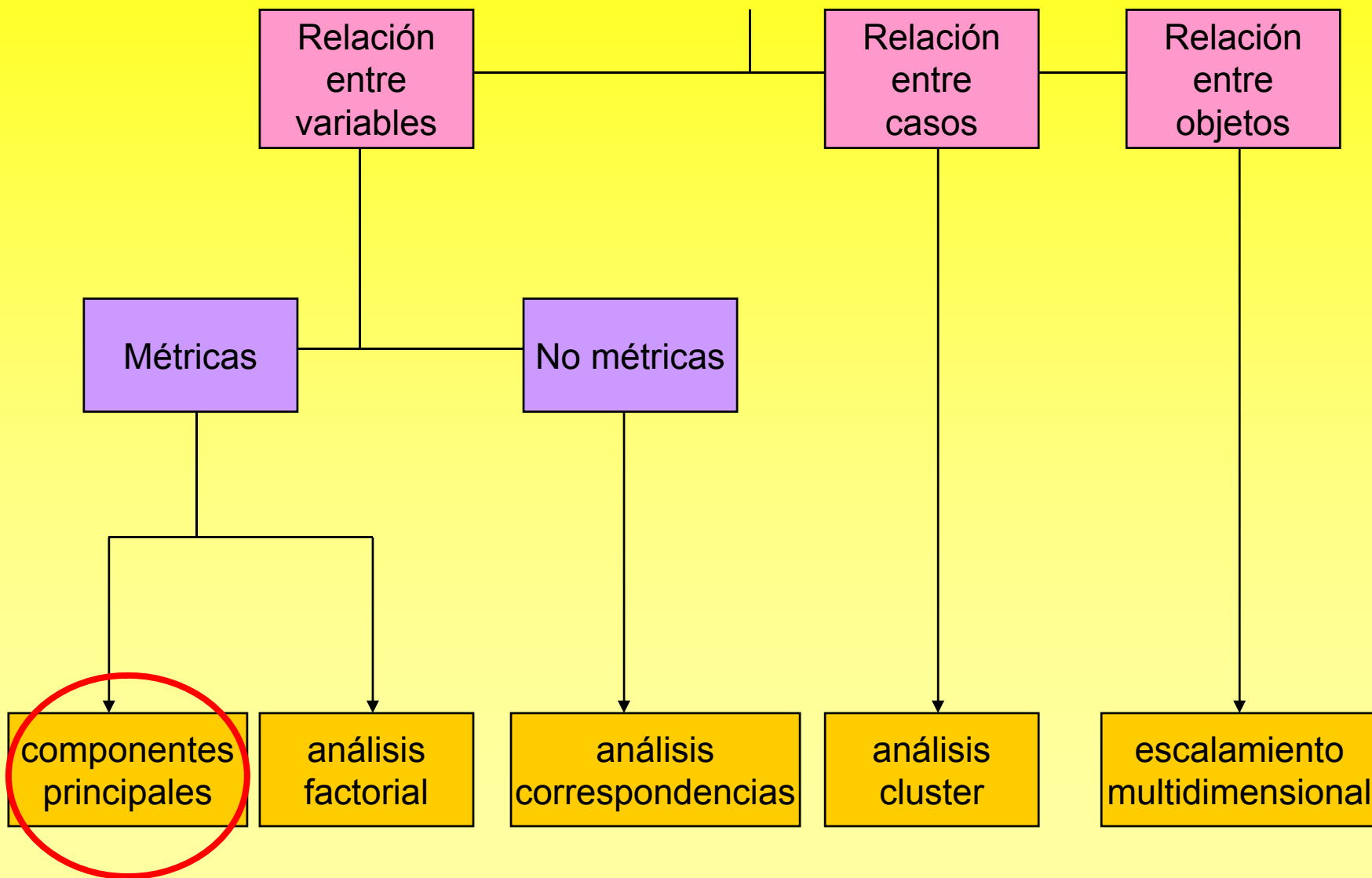
ANÁLISIS DE COMPONENTES PRINCIPALES



COMPONENTES PRINCIPALES

1. Introducción
2. Componentes principales de dos variables
3. Obtención de los componentes principales
4. Caso práctico

Análisis de interdependencias



Objetivo. Transformar un conjunto de variables (***variables originales***) en un nuevo conjunto de variables (***componentes principales***), **incorrelacionadas** entre sí.

Objetivo. Transformar un conjunto de variables (*variables originales*) en un nuevo conjunto de variables (*componentes principales*), **incorrelacionadas entre sí.**

Interés.

- Para explicar fenómenos cuya información se cifra en **muchas variables** más o menos **correlacionadas**.
- Reducir la **dimensión** del número de variables inicialmente consideradas en el análisis.
- Las nuevas variables pueden **ordenarse** según la información que llevan.

Principios básicos.

- Sólo con **datos cuantitativos** y no es necesario establecer jerarquías ni comprobar la normalidad.

Principios básicos.

- Sólo con **datos cuantitativos** y no es necesario establecer jerarquías ni comprobar la normalidad.
- Si las *variables originales* no están **correlacionadas**, el análisis no tiene sentido.

Principios básicos.

- Sólo con **datos cuantitativos** y no es necesario establecer jerarquías ni comprobar la normalidad.
- Si las *variables originales* no están **correlacionadas**, el análisis no tiene sentido.
- Como medida de la cantidad de **información** incorporada en el *componente* se utiliza la **varianza (6)**.

Principios básicos.

- Sólo con **datos cuantitativos** y no es necesario establecer jerarquías ni comprobar la normalidad.
- Si las *variables originales* no están **correlacionadas**, el análisis no tiene sentido.
- Como medida de la cantidad de **información** incorporada en el *componente* se utiliza la **varianza (6)**.
- Por tanto se **ordenan** según la varianza de mayor a menor.

Principios básicos.

- Se trabaja con **variables tipificadas** o con **variables expresadas en desviaciones respecto a la media** para evitar problemas derivados de la **escala**.

Principios básicos.

- Se trabaja con **variables tipificadas** o con **variables expresadas en desviaciones respecto a la media** para evitar problemas derivados de la **escala**.
- El nuevo conjunto de variables (*componentes principales*) es igual al número de *variables originales*.

Principios básicos.

- Se trabaja con **variables tipificadas** o con **variables expresadas en desviaciones respecto a la media** para evitar problemas derivados de la **escala**.
- El nuevo conjunto de variables (*componentes principales*) es igual al número de *variables originales*.
- Los componentes principales se expresan como una combinación lineal de las variables originales.

Utilidad. Si las variables que explican un fenómeno son muchas y están correlacionadas, es posible explicar el fenómeno con muy pocos componentes principales.

- Análisis de mercados.
- Determinación del proceso de compra.
- Caracterización de explotaciones.
- etc.

Componentes principales de 2 variables

Ejemplo: Estudiar el beneficio y la dimensión de 9 explotaciones bovinas ecológicas

¿Qué variables vamos a utilizar?

Componentes principales de 2 variables

Ejemplo: Estudiar el beneficio y la dimensión de 9 explotaciones bovinas ecológicas

¿Qué variables vamos a utilizar?

Beneficio: € por explotación

Dimensión: inversión € por explotación

Componentes principales de 2 variables

	variables originales	
explotación	inversión (€)	beneficios (€)
1	775.104	23.795
2	775.218	58.778
3	700.963	1.531
4	674.063	-12.756
5	631.003	14.729
6	537.744	9.059
7	489.155	12.541
8	448.465	13.495
9	445.853	-34.828

Componentes principales de 2 variables

	variables originales	
explotación	inversión (€)	beneficios (€)
1	775.104	23.795
2	775.218	58.778
3	700.963	1.531
4	674.063	-12.756
5	631.003	14.729
6	537.744	9.059
7	489.155	12.541
8	448.465	13.495
9	445.853	-34.828

Primer paso:

¿Hay correlación entre ambas variables?

Componentes principales de 2 variables

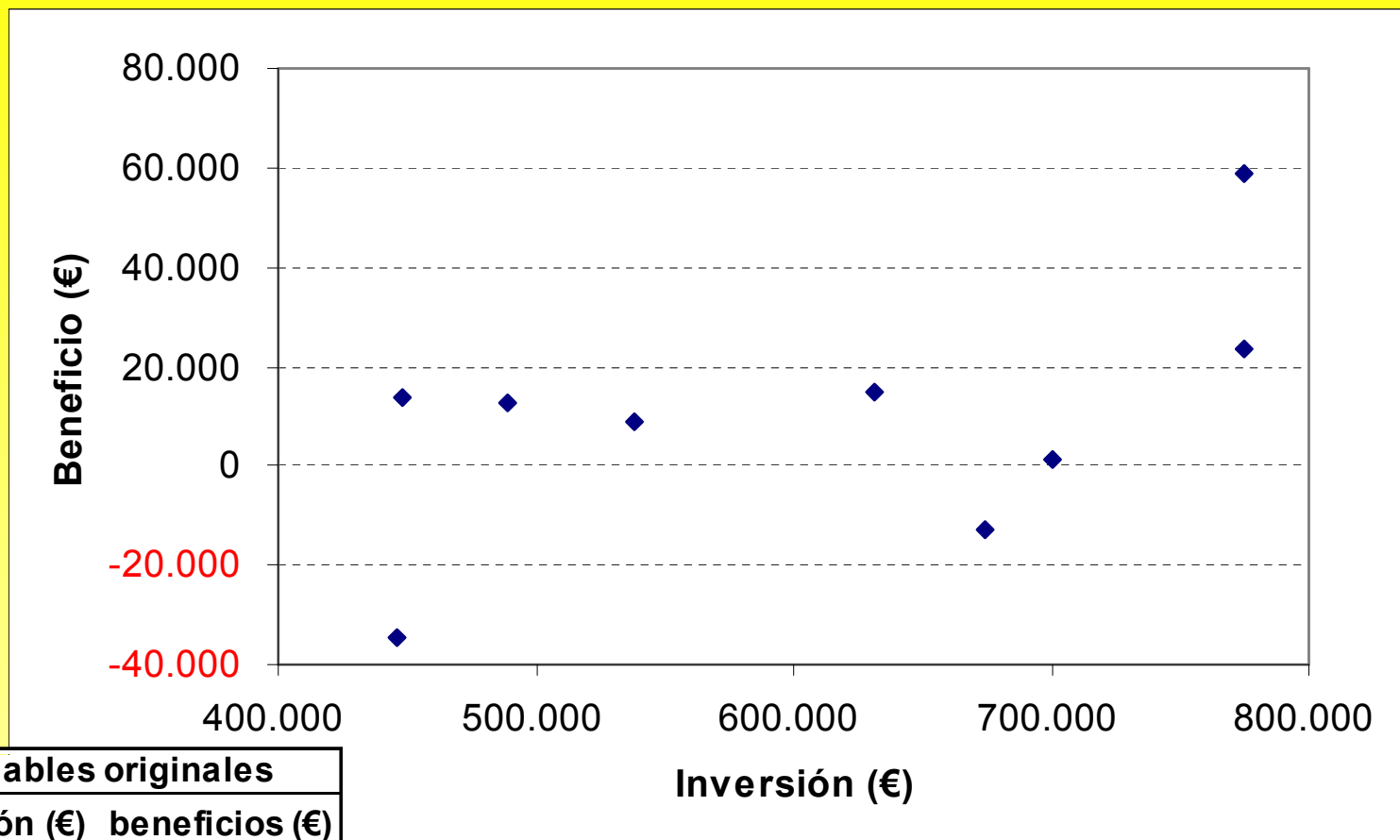
```
Correlations
          n_vacas      beneficio
-----
n_vacas          0,5460
                ( 9)
                0,1283

beneficio      0,5460
              ( 9)
              0,1283
-----
```

Primer paso:

¿Hay correlación entre ambas variables?

Componentes principales de 2 variables



	variables originales	
explotación	inversión (€)	beneficios (€)
1	775.104	23.795
2	775.218	58.778
3	700.963	1.531
4	674.063	-12.756
5	631.003	14.729
6	537.744	9.059
7	489.155	12.541
8	448.465	13.495
9	445.853	-34.828

Componentes principales de 2 variables

	variables originales	
explotación	inversión (€)	beneficios (€)
1	775.104	23.795
2	775.218	58.778
3	700.963	1.531
4	674.063	-12.756
5	631.003	14.729
6	537.744	9.059
7	489.155	12.541
8	448.465	13.495
9	445.853	-34.828

Segundo paso:

Eliminar el problema de la escala

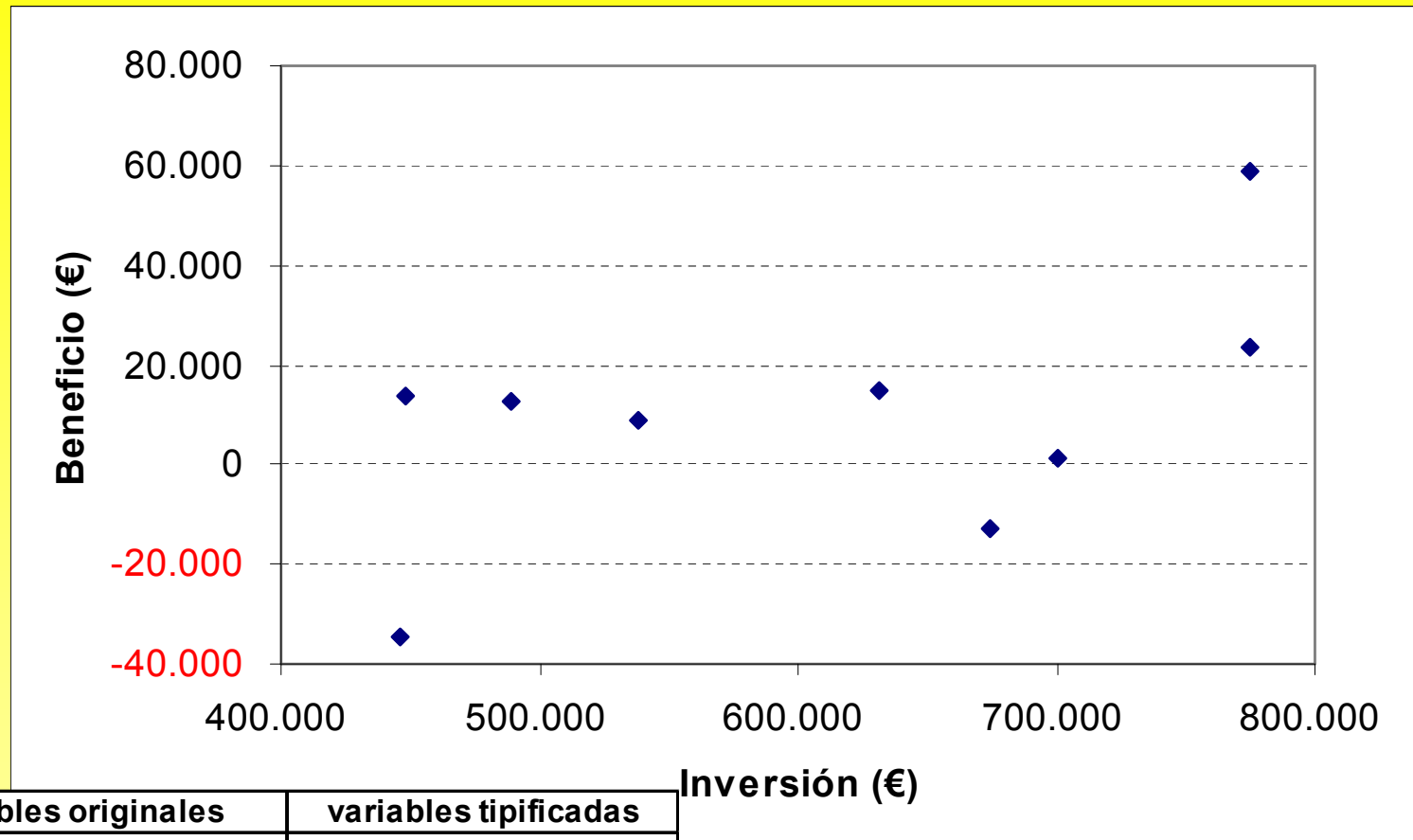
Componentes principales de 2 variables

	variables originales		variables tipificadas	
explotación	inversión (€)	beneficios (€)	inversión	beneficios
1	775.104	23.795	1,257	0,556
2	775.218	58.778	1,258	1,927
3	700.963	1.531	0,697	-0,316
4	674.063	-12.756	0,494	-0,875
5	631.003	14.729	0,169	0,201
6	537.744	9.059	-0,535	-0,020
7	489.155	12.541	-0,902	0,115
8	448.465	13.495	-1,209	0,152
9	445.853	-34.828	-1,229	-1,749

Tipificar las variables

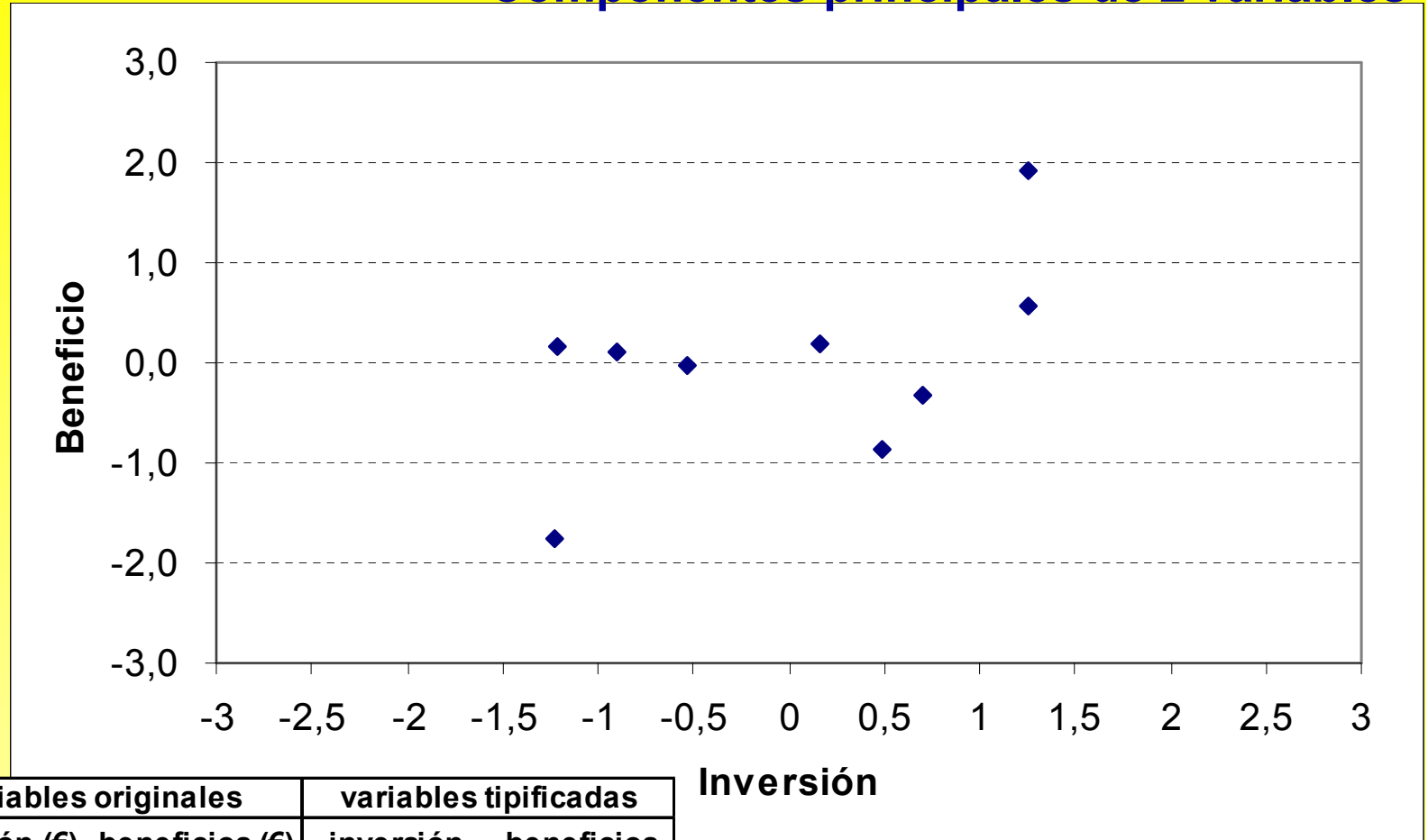
- La **matriz de correlación** es igual a la matriz de covarianzas
- $\sum \sigma_{\text{componentes principales}} = \sum \sigma_{\text{variables}} = \sum \text{variables tipificadas (2 en este caso)}$

Componentes principales de 2 variables



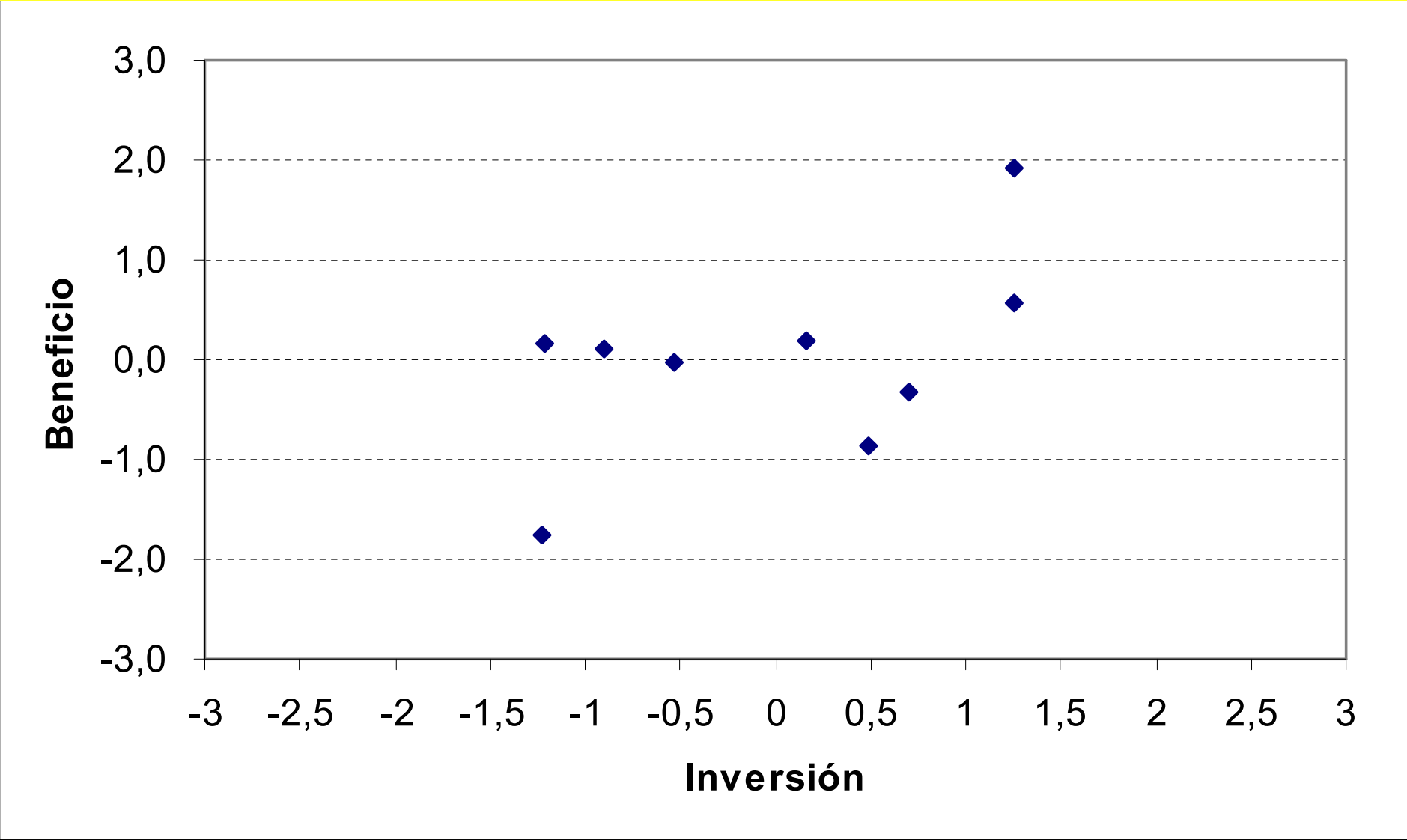
explotación	variables originales		variables tipificadas	
	inversión (€)	beneficios (€)	inversión	beneficios
1	775.104	23.795	1,257	0,556
2	775.218	58.778	1,258	1,927
3	700.963	1.531	0,697	-0,316
4	674.063	-12.756	0,494	-0,875
5	631.003	14.729	0,169	0,201
6	537.744	9.059	-0,535	-0,020
7	489.155	12.541	-0,902	0,115
8	448.465	13.495	-1,209	0,152
9	445.853	-34.828	-1,229	-1,749

Componentes principales de 2 variables

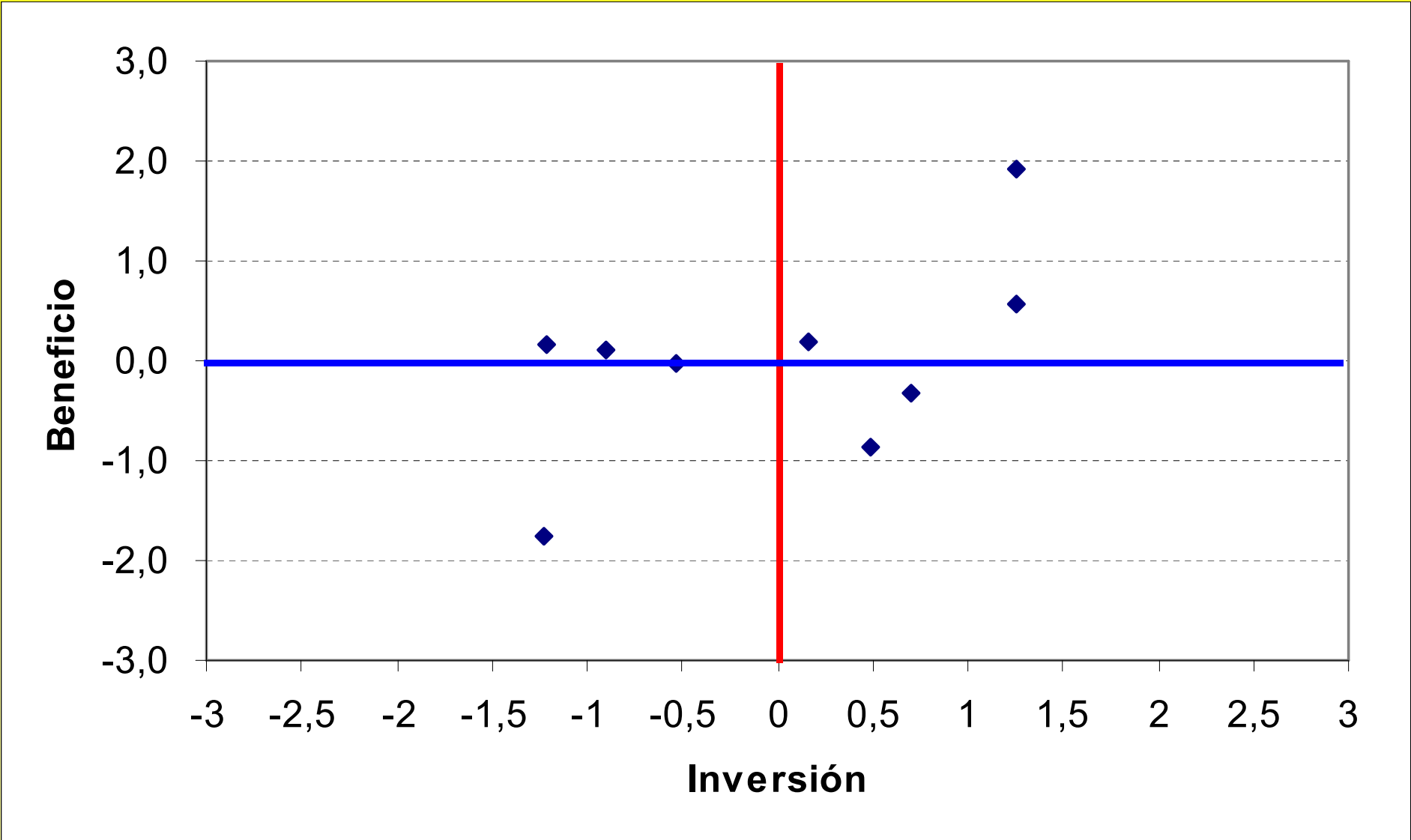


	variables originales		variables tipificadas	
explotación	inversión (€)	beneficios (€)	inversión	beneficios
1	775.104	23.795	1,257	0,556
2	775.218	58.778	1,258	1,927
3	700.963	1.531	0,697	-0,316
4	674.063	-12.756	0,494	-0,875
5	631.003	14.729	0,169	0,201
6	537.744	9.059	-0,535	-0,020
7	489.155	12.541	-0,902	0,115
8	448.465	13.495	-1,209	0,152
9	445.853	-34.828	-1,229	-1,749

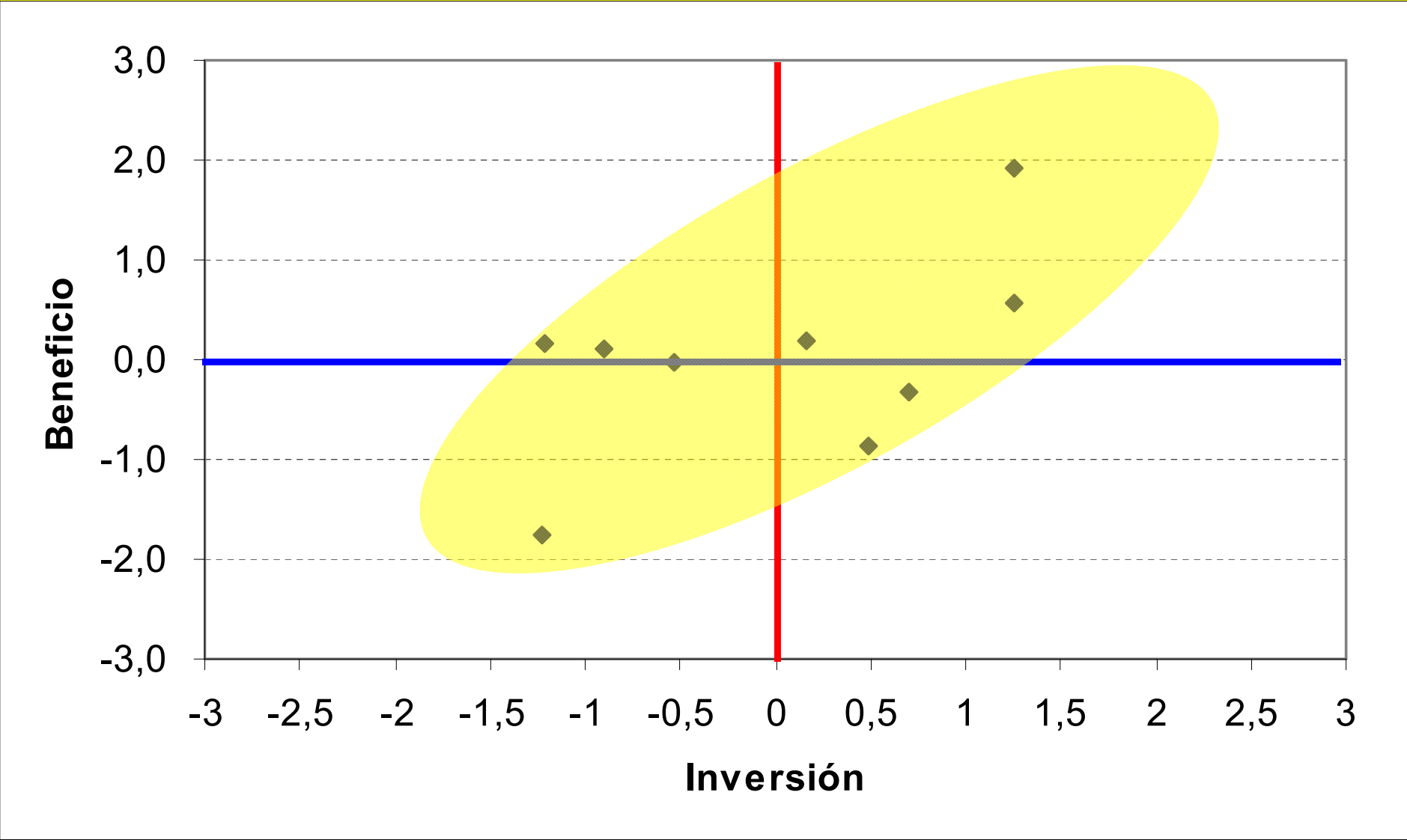
Componentes principales de 2 variables



Componentes principales de 2 variables

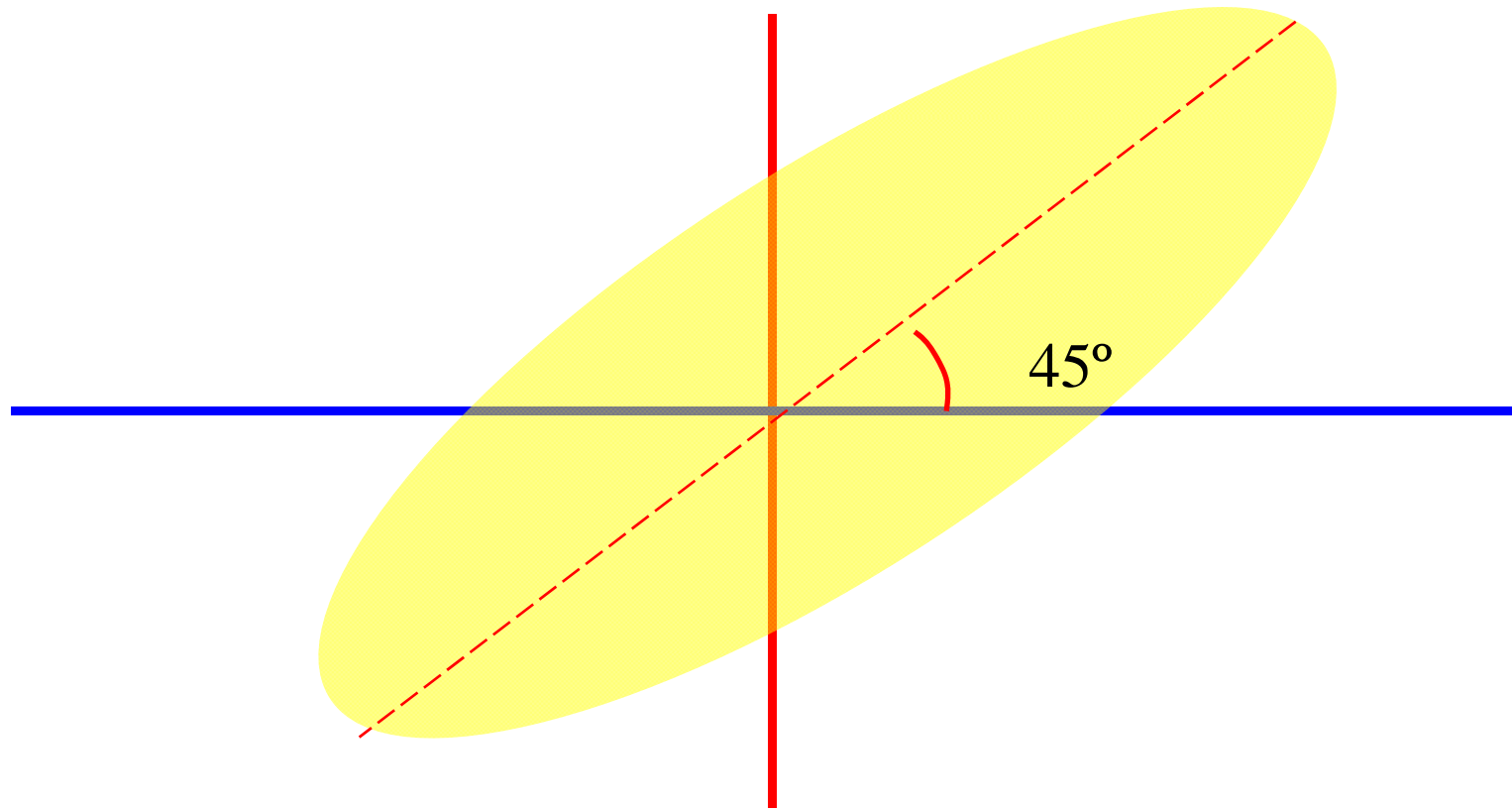


Componentes principales de 2 variables



Elipse de concentración

Componentes principales de 2 variables



Elipse de concentración en una distribución normal bivalente

Componentes principales de 2 variables

	variables originales	
explotación	inversión (€)	beneficios (€)
1	775.104	23.795
2	775.218	58.778
3	700.963	1.531
4	674.063	-12.756
5	631.003	14.729
6	537.744	9.059
7	489.155	12.541
8	448.465	13.495
9	445.853	-34.828

Tercer paso:

Obtener los componentes principales

Componentes principales de 2 variables

1. Calcular las **raíces** de la matriz de covarianzas:

Componentes principales de 2 variables

1. Calcular las raíces de la matriz de covarianzas:

$$\lambda_1 = 1,54603$$

$$\lambda_2 = 0,45397$$

Componentes principales de 2 variables

1. Calcular las **raíces** de la matriz de covarianzas:

$$\lambda_1 = 1,54603$$

$$\lambda_2 = 0,45397$$

- La σ de cada CP es igual al valor de la **raíz característica**.

Componentes principales de 2 variables

1. Calcular las **raíces** de la matriz de covarianzas:

$$\lambda_1 = 1,54603$$

$$\lambda_2 = 0,45397$$

- La σ de cada CP es igual al valor de la **raíz característica**.
- La primera CP se obtiene de forma que maximice σ .
 - En general tiene una **σ mayor** que cualquier variable original.
 - Si la variable está tipificada, la **σ CP1 > 1**.

Componentes principales de 2 variables

1. Calcular las raíces de la matriz de covarianzas:

$$\lambda_1 = 1,54603$$

$$\lambda_2 = 0,45397$$

- La σ de cada CP es igual al valor de la **raíz característica**.
- La primera CP se obtiene de forma que maximice σ .
 - En general tiene una **σ mayor** que cualquier variable original.
 - Si la variable está tipificada, la **σ CP1 > 1**.
- Si las variables originales están incorrelacionadas, las CP coincidirán exactamente con las variables originales.

Componentes principales de 2 variables

- En el caso de 2 variables, $\hat{\beta}$ de CP1 = $\hat{\beta}$ de una de las variables tipificadas + coeficiente de correlación:

Componentes principales de 2 variables

- En el caso de 2 variables, $\hat{\sigma}$ de CP1 = $\hat{\sigma}$ de una de las variables tipificadas + coeficiente de correlación: $1 + 0,54603 = 1,54603$

$$\lambda_1 = 1,54603$$

$$\lambda_2 = 0,45397$$

Componentes principales de 2 variables

- En el caso de 2 variables, $\hat{\sigma}$ de CP1 = $\hat{\sigma}$ de una de las variables tipificadas + coeficiente de correlación: $1 + 0,54603 = 1,54603$

$$\lambda_1 = 1,54603$$

$$\lambda_2 = 0,45397$$

- $\hat{\sigma}$ de CP2 = $\sqrt{2}$ - $\hat{\sigma}$ de CP1

Componentes principales de 2 variables

- En el caso de 2 variables, σ de CP1 = σ de una de las variables tipificadas + coeficiente de correlación: $1 + 0,54603 = 1,54603$

$$\lambda_1 = 1,54603$$

$$\lambda_2 = 0,45397$$

- σ de CP2 = $\textcircled{2}$ - σ de CP1

$$\sum \sigma_{\text{componentes principales}} = \sum \sigma_{\text{variables}} = \sum \text{variables tipificadas}$$

Componentes principales de 2 variables

Con el Statgrafics:

Analysis Summary

Data variables:

inversion

beneficio

Data input: observations

Number of complete cases: 9

Missing value treatment: listwise

Standardized: yes

Number of components extracted: 2

Principal Components Analysis

Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	1,54604	77,302	77,302
2	0,453963	22,698	100,000

Componentes principales de 2 variables

2. Cada raíz tiene asociado un **vector** característico, que con dos variables:

$$u_1 = \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} \quad y \quad u_2 = \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix}$$

Componentes principales de 2 variables

2. Cada raíz tiene asociado un **vector** característico, que con dos variables:

$$u_1 = \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} \quad y \quad u_2 = \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix}$$

- Deben cumplir: $u_{11}^2 + u_{12}^2 = 1$
 $u_{21}^2 + u_{22}^2 = 1$

Componentes principales de 2 variables

2. Cada raíz tiene asociado un **vector** característico, que con dos variables:

$$u_1 = \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} \quad y \quad u_2 = \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix}$$

- Deben cumplir: $u_{11}^2 + u_{12}^2 = 1$

$$u_{21}^2 + u_{22}^2 = 1$$

- Si los datos están tipificados, siempre con 2 variables se obtienen los siguientes vectores:

$$u_1 = \begin{bmatrix} 0,7071 \\ 0,7071 \end{bmatrix} \quad y \quad u_2 = \begin{bmatrix} 0,7071 \\ -0,7071 \end{bmatrix}$$

Componentes principales de 2 variables

- Los coeficientes de los vectores son los coeficientes que hay que aplicar a las variables **tipificadas** para obtener los CP:

$$CP1 = u_{11} * X_1 + u_{12} * X_2$$

$$CP2 = u_{21} * X_1 + u_{22} * X_2$$

Componentes principales de 2 variables

- Los coeficientes de los vectores son los coeficientes que hay que aplicar a las variables **tipificadas** para obtener los CP:

$$CP1 = u_{11} * X_1 + u_{12} * X_2$$

$$CP2 = u_{21} * X_1 + u_{22} * X_2$$

- En nuestro caso:

$$u_1 = \begin{bmatrix} 0,7071 \\ 0,7071 \end{bmatrix} \text{ y } u_2 = \begin{bmatrix} 0,7071 \\ -0,7071 \end{bmatrix}$$

Componentes principales de 2 variables

- Los coeficientes de los vectores son los coeficientes que hay que aplicar a las variables **tipificadas** para obtener los CP:

$$CP1 = u_{11} * X_1 + u_{12} * X_2$$

$$CP2 = u_{21} * X_1 + u_{22} * X_2$$

- En nuestro caso:

$$u_1 = \begin{bmatrix} 0,7071 \\ 0,7071 \end{bmatrix} \text{ y } u_2 = \begin{bmatrix} 0,7071 \\ -0,7071 \end{bmatrix}$$

$$CP1 = 0,7071 * \text{inversión} + 0,7071 * \text{beneficio}$$

$$CP2 = 0,7071 * \text{inversión} - 0,7071 * \text{beneficio}$$

Componentes principales de 2 variables

Con el Statgrafics:

Table of Component Weights

	Component 1	Component 2
	-----	-----
inversion	0,707107	0,707107
beneficio	0,707107	-0,707107

The StatAdvisor

This table shows the equations of the principal components. For example, the first principal component has the equation

$$0,707107 * inversion + 0,707107 * beneficio$$

Componentes principales de 2 variables

$$CP1 = 0,7071 * inversión + 0,7071 * beneficio$$

$$CP2 = 0,7071 * inversión - 0,7071 * beneficio$$

¿Qué significado tienen los coeficientes?

Componentes principales de 2 variables

$$CP1 = 0,7071 * inversión + 0,7071 * beneficio$$

$$CP2 = 0,7071 * inversión - 0,7071 * beneficio$$

¿Qué significado tienen los coeficientes?

Son los senos y los cosenos del ángulo de rotación entre los ejes de los CP y los ejes de las variables tipificadas

Componentes principales de 2 variables

$$\text{CP1} = 0,7071 * \text{inversión} + 0,7071 * \text{beneficio}$$

$$\text{CP2} = 0,7071 * \text{inversión} - 0,7071 * \text{beneficio}$$

Componentes principales de 2 variables

$$\text{CP1} = 0,7071 * \text{inversión} + 0,7071 * \text{beneficio}$$

Primer eje: $\cos 45^\circ = 0,7071$ $\text{sen } 45^\circ = 0,7071$

$$\text{CP2} = 0,7071 * \text{inversión} - 0,7071 * \text{beneficio}$$

Componentes principales de 2 variables

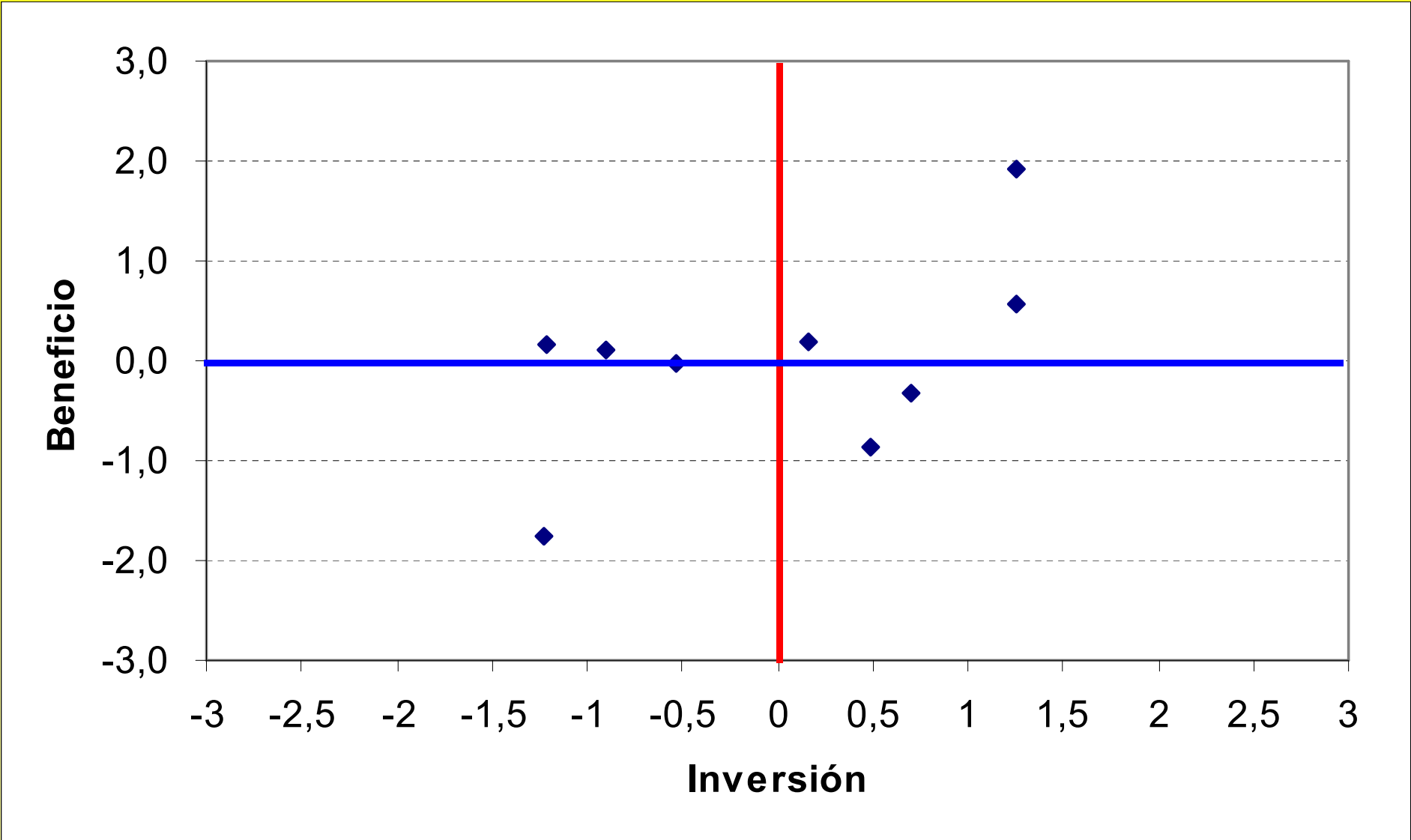
$$\text{CP1} = 0,7071 * \text{inversión} + 0,7071 * \text{beneficio}$$

$$\text{Primer eje:} \quad \cos 45^\circ = 0,7071 \quad \text{sen } 45^\circ = 0,7071$$

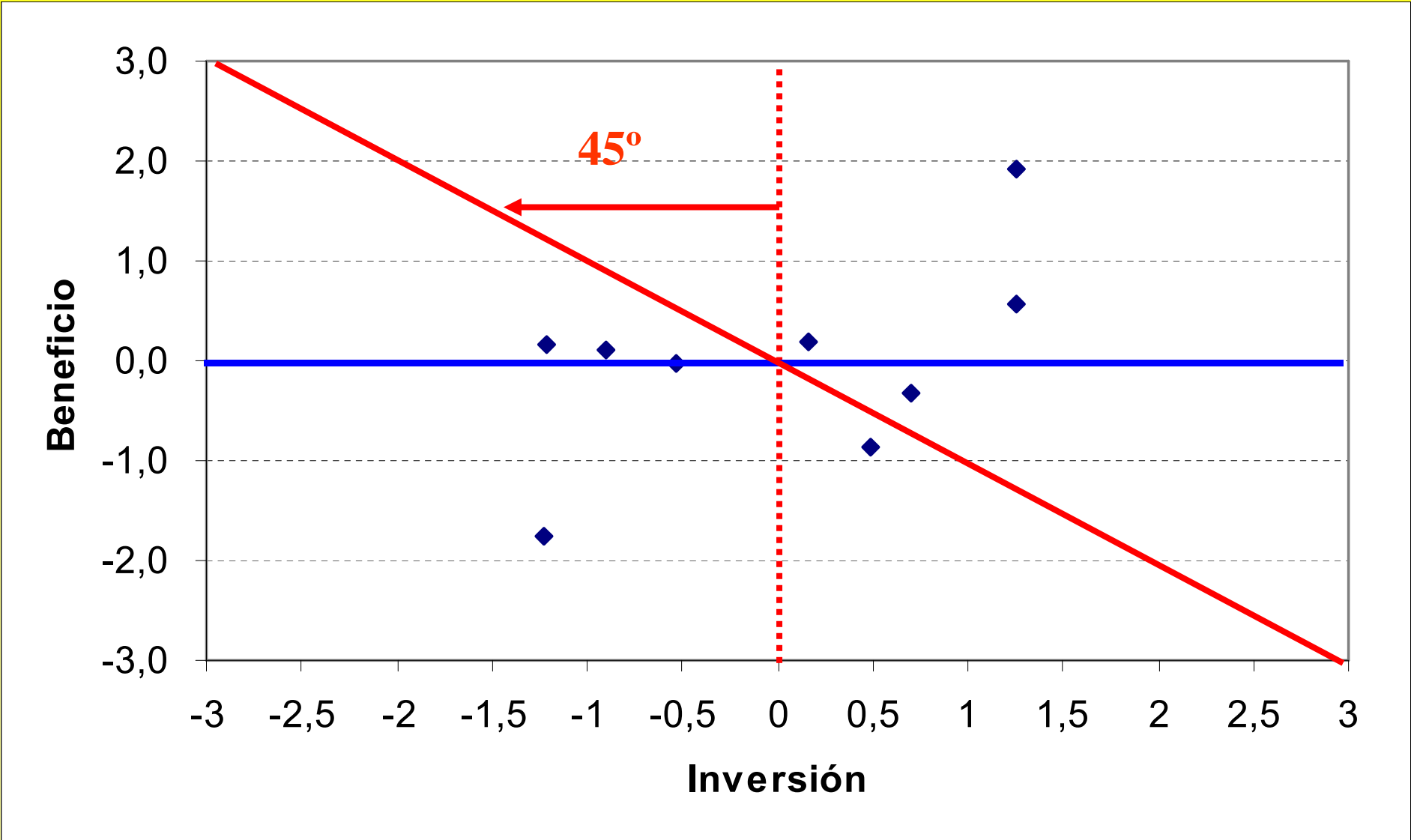
$$\text{CP2} = 0,7071 * \text{inversión} - 0,7071 * \text{beneficio}$$

$$\text{Segundo eje:} \quad \cos 135^\circ = 0,7071 \quad \text{sen } 135^\circ = -0,7071$$

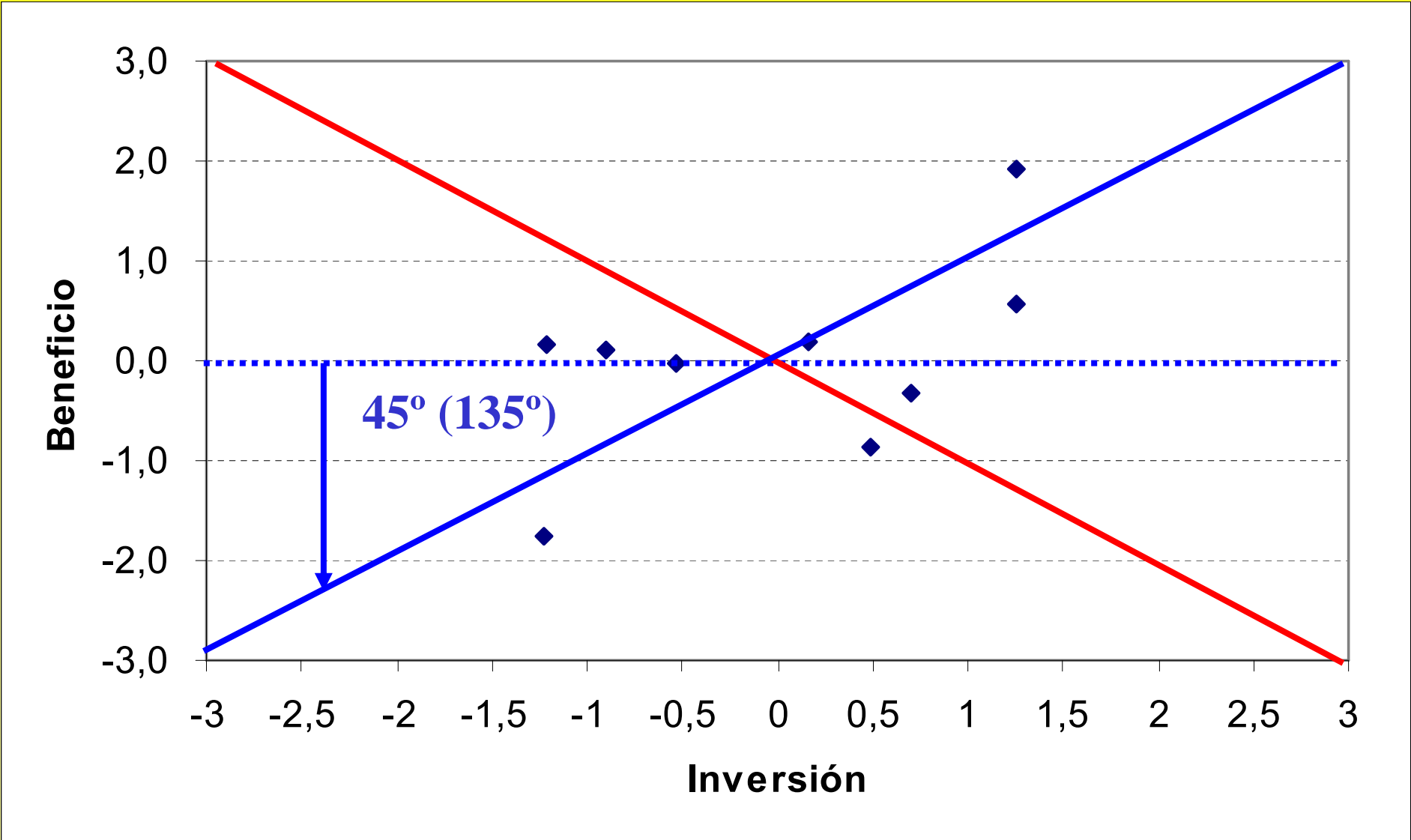
Componentes principales de 2 variables



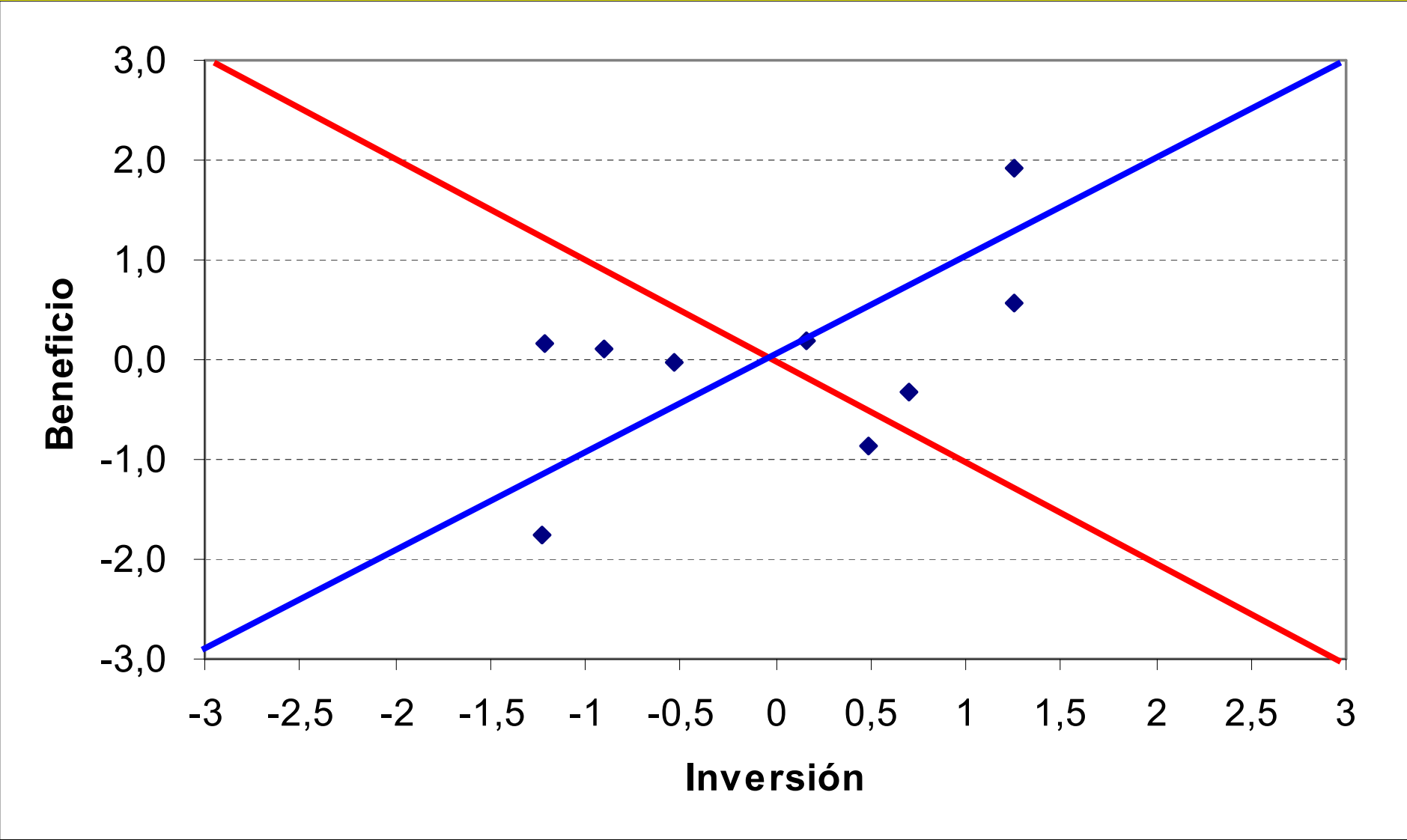
Componentes principales de 2 variables



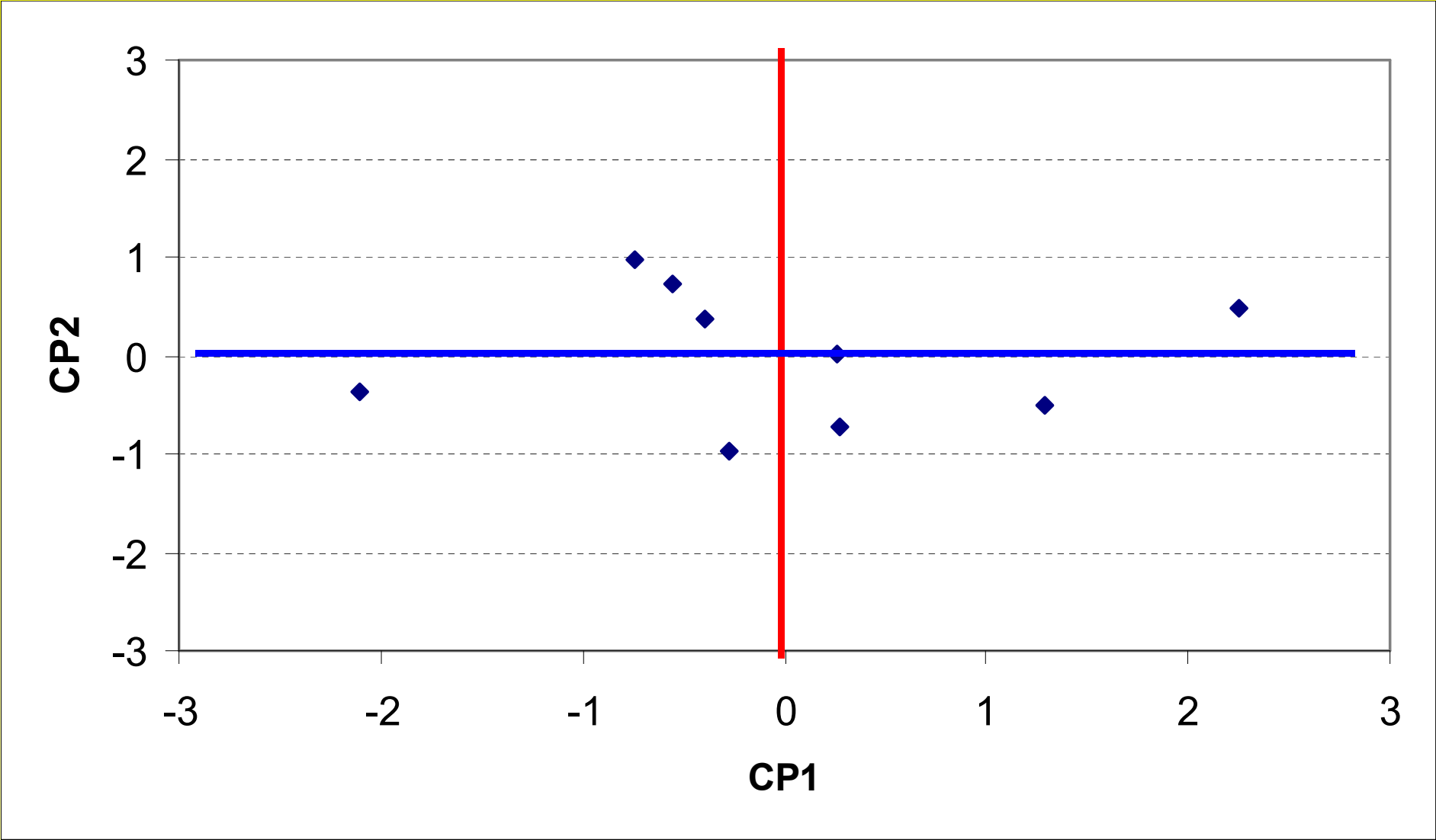
Componentes principales de 2 variables



Componentes principales de 2 variables



Componentes principales de 2 variables



Componentes principales de 2 variables

3. Determinar las **cargas factoriales**:

Componentes principales de 2 variables

3. Determinar las **cargas factoriales**:

- Correlación de cada variables con cada CP

Componentes principales de 2 variables

3. Determinar las **cargas factoriales**:

- Correlación de cada variables con cada CP
- Coeficiente de correlación r_{hj} entre el componente h y la variable j:

$$r_{hj} = u_{hj} * \sqrt{\lambda_h}$$

Componentes principales de 2 variables

3. Determinar las **cargas factoriales**:

- Correlación de cada variables con cada CP
- Coeficiente de correlación r_{hj} entre el componente h y la variable j:

$$r_{hj} = u_{hj} * \sqrt{\lambda_h}$$

- En nuestro caso (matriz factorial o matriz de componentes):

$$\text{CP1 con inversión: } 0,7071 * \sqrt{1,54603} = 0,87821$$

$$\text{CP1 con beneficio: } 0,7071 * \sqrt{1,54603} = 0,87821$$

$$\text{CP2 con inversión: } 0,7071 * \sqrt{0,43397} = 0,47643$$

$$\text{CP1 con beneficio: } -0,7071 * \sqrt{0,43397} = -0,47643$$

Componentes principales de 2 variables

Con el Statgrafics:

Correlations

	inversion	beneficio	PCOMP_1
inversion		0,5460 (9) 0,1283	0,8792 (9) 0,0018
beneficio	0,5460 (9) 0,1283		0,8792 (9) 0,0018
PCOMP_1	0,8792 (9) 0,0018	0,8792 (9) 0,0018	
PCOMP_2	0,4764 (9) 0,1948	-0,4764 (9) 0,1948	0,0000 (9) 1,0000

	PCOMP_2
inversion	0,4764 (9) 0,1948
beneficio	-0,4764 (9) 0,1948
PCOMP_1	0,0000 (9) 1,0000

Componentes principales de 2 variables

4. Determinar las puntuaciones tipificadas de cada componente:

$$CP1 = 0,7071 * inversión + 0,7071 * beneficio$$

$$CP2 = 0,7071 * inversión - 0,7071 * beneficio$$

teniendo en cuenta que inversión y beneficio están tipificados (- X / D.E.)



Componentes principales de 2 variables

Con el Statgrafics:

Table of Principal Components

Row	Component 1	Component 2
1	1,28273	0,495675
2	2,25274	-0,473118
3	0,269787	0,716637
4	-0,269788	0,968866
5	0,261855	-0,0227454
6	-0,393362	-0,363724
7	-0,556388	-0,719727
8	-0,747277	-0,963488
9	-2,10029	0,361623

Componentes principales de 2 variables

	variables originales		variables tipificadas		componentes principales	
explotación	inversión (€)	beneficios (€)	inversión	beneficios	CP1	CP2
1	775.104	23.795	1,257	0,556	1,283	-0,496
2	775.218	58.778	1,258	1,927	2,253	0,473
3	700.963	1.531	0,697	-0,316	0,270	-0,717
4	674.063	-12.756	0,494	-0,875	-0,270	-0,969
5	631.003	14.729	0,169	0,201	0,262	0,023
6	537.744	9.059	-0,535	-0,020	-0,393	0,364
7	489.155	12.541	-0,902	0,115	-0,556	0,720
8	448.465	13.495	-1,209	0,152	-0,747	0,963
9	445.853	-34.828	-1,229	-1,749	-2,100	-0,362

Tercer paso:

Obtener los componentes principales

- Es un caso típico de cálculo de raíces y vectores en matrices simétricas.

- Es un caso típico de cálculo de raíces y vectores en matrices simétricas.
- Los CP sin combinaciones lineales de las variables originales:

$$CP1 = 0,7071 * inversión + 0,7071 * beneficio$$

$$CP2 = 0,7071 * inversión - 0,7071 * beneficio$$

- Es un caso típico de cálculo de raíces y vectores en matrices simétricas.
- Los CP son combinaciones lineales de las variables originales.
- Los coeficientes de las combinaciones lineales son los elementos característicos asociados a la matriz de covarianzas de las variables originales.

$$CP1 = u_{11} * inversión + u_{12} * beneficio$$

$$CP2 = u_{21} * inversión + u_{22} * beneficio$$

- Es un caso típico de cálculo de raíces y vectores en matrices simétricas.
- Los CP son combinaciones lineales de las variables originales.
- Los coeficientes de las combinaciones lineales son los elementos característicos asociados a la matriz de covarianzas de las variables originales.
- La σ de cada CP es igual a la raíz característica a que va asociada.

$$\lambda_1 = 1,54603$$

$$\lambda_2 = 0,45397$$

- Es un caso típico de cálculo de raíces y vectores en matrices simétricas.
- Los CP son combinaciones lineales de las variables originales.
- Los coeficientes de las combinaciones lineales son los elementos característicos asociados a la matriz de covarianzas de las variables originales.
- La σ de cada CP es igual a la raíz característica a que va asociada.
- Si las variables están tipificadas, la proporción de variabilidad original captada por un CP es igual a su raíz característica dividida por el número de variables originales.

Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	1,54604	77,302	77,302
2	0,453963	22,698	100,000

- Es un caso típico de cálculo de raíces y vectores en matrices simétricas.
- Los CP son combinaciones lineales de las variables originales.
- Los coeficientes de las combinaciones lineales son los elementos característicos asociados a la matriz de covarianzas de las variables originales.
- La σ de cada CP es igual a la raíz característica a que va asociada.
- Si las variables están tipificadas, la proporción de variabilidad original captada por un CP es igual a su raíz característica dividida por el número de variables originales.
- Si las variables están tipificadas, la correlación entre cada CP con cada variable se determina con la raíz característica de la CP y el elemento del del vector característico asociado.

$$r_{hj} = u_{hj} * \sqrt{\lambda_h}$$

1. Obtención de la primera componente.

- Muestra de tamaño n sobre las siguientes variables. X_1, X_2, \dots, X_p .

1. Obtención de la primera componente.

- Muestra de tamaño n sobre las siguientes variables. X_1, X_2, \dots, X_p .
- Las observaciones están tipificadas.

1. Obtención de la primera componente.

- Muestra de tamaño n sobre las siguientes variables. X_1, X_2, \dots, X_p .
- Las observaciones están tipificadas.
- La primera componente sería:

$$CP_{1i} = u_{11} * X_{1i} + u_{12} * X_{2i} + \dots + u_{1p} * X_{pi}$$

1. Obtención de la primera componente.

- Muestra de tamaño n sobre las siguientes variables. X_1, X_2, \dots, X_p .
- Las observaciones están tipificadas.
- La primera componente sería:

$$CP_{1i} = u_{11} * X_{1i} + u_{12} * X_{2i} + \dots + u_{1p} * X_{pi}$$

- Para el conjunto de las n observaciones:

$$\begin{pmatrix} CP_{11} \\ CP_{12} \\ \dots \\ CP_{1n} \end{pmatrix} = \begin{pmatrix} X_{11} & X_{21} & X_{31} & \dots & X_{p1} \\ X_{12} & X_{22} & X_{32} & \dots & X_{p2} \\ \dots & \dots & \dots & \dots & \dots \\ X_{1n} & X_{2n} & X_{3n} & \dots & X_{pn} \end{pmatrix} \begin{pmatrix} u_{11} \\ u_{12} \\ \dots \\ u_{1p} \end{pmatrix}$$

1. Obtención de la primera componente.

- Se busca que la primera componente tenga varianza máxima, considerando que la suma de los pesos (u_{1j}) al cuadrado sea igual a la unidad.

1. Obtención de la primera componente.

- Se busca que la primera componente tenga varianza máxima, considerando que la suma de los pesos (u_{1j}) al cuadrado sea igual a la unidad.
- La varianza a maximizar sería $u_1'Vu_1$ (siendo V la matriz de correlaciones).

1. Obtención de la primera componente.

- Se busca que la primera componente tenga varianza máxima, considerando que la suma de los pesos (u_{1j}) al cuadrado sea igual a la unidad.
- La varianza a maximizar sería $u_1'Vu_1$ (siendo V la matriz de correlaciones).
- Al resolver la ecuación (primera derivada de u_1 e igualar a 0): $(V - \lambda I)u_1 = 0$, se obtienen p raíces características y se toma la mayor (λ_1).
- Con la mayor λ_1 se obtiene el vector asociado u_1 que corresponde al vector característico asociado a la raíz característica mayor de la matriz V .

2. Obtención de los restantes componentes.

- Se repite el mismo proceso aunque se impone la restricción de que cada nuevo vector sea ortogonal a los anteriores:

$$u_h^T u_1 = u_h^T u_2 = \dots = u_h^T u_{h-1} = 0$$

2. Obtención de los restantes componentes.

- Se repite el mismo proceso aunque se impone la restricción de que cada nuevo vector sea ortogonal a los anteriores:

$$u_h u_1 = u_h u_2 = \dots = u_h u_{h-1} = 0$$

- Por tanto, las p CP que se pueden calcular siempre son combinación lineal de las variables originales y los coeficientes de ponderación son los correspondientes vectores característicos asociados a la matriz V (matriz de covarianzas).

3. Número de componentes a retener.

3. Número de componentes a retener.

- **Criterio de la media aritmética.** Se seleccionan aquellas CP cuya raíz característica (λ) supere la media de las raíces características.
 - Si tenemos variables tipificadas, todas aquellas que superen el valor 1.

Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	1,54604	77,302	77,302
2	0,453963	22,698	100,000

3. Número de componentes a retener.

- **Criterio de la media aritmética.** Se seleccionan aquellas CP cuya raíz característica (λ) supere la media de las raíces características.
 - Si tenemos variables tipificadas, todas aquellas que superen

```
Data variables:  
  inversion  
  beneficio  
  vacas  
  hectareas
```

```
Data input: observations  
Number of complete cases: 9  
Missing value treatment: listwise  
Standardized: yes
```

```
Number of components extracted: 4
```

Principal Components Analysis

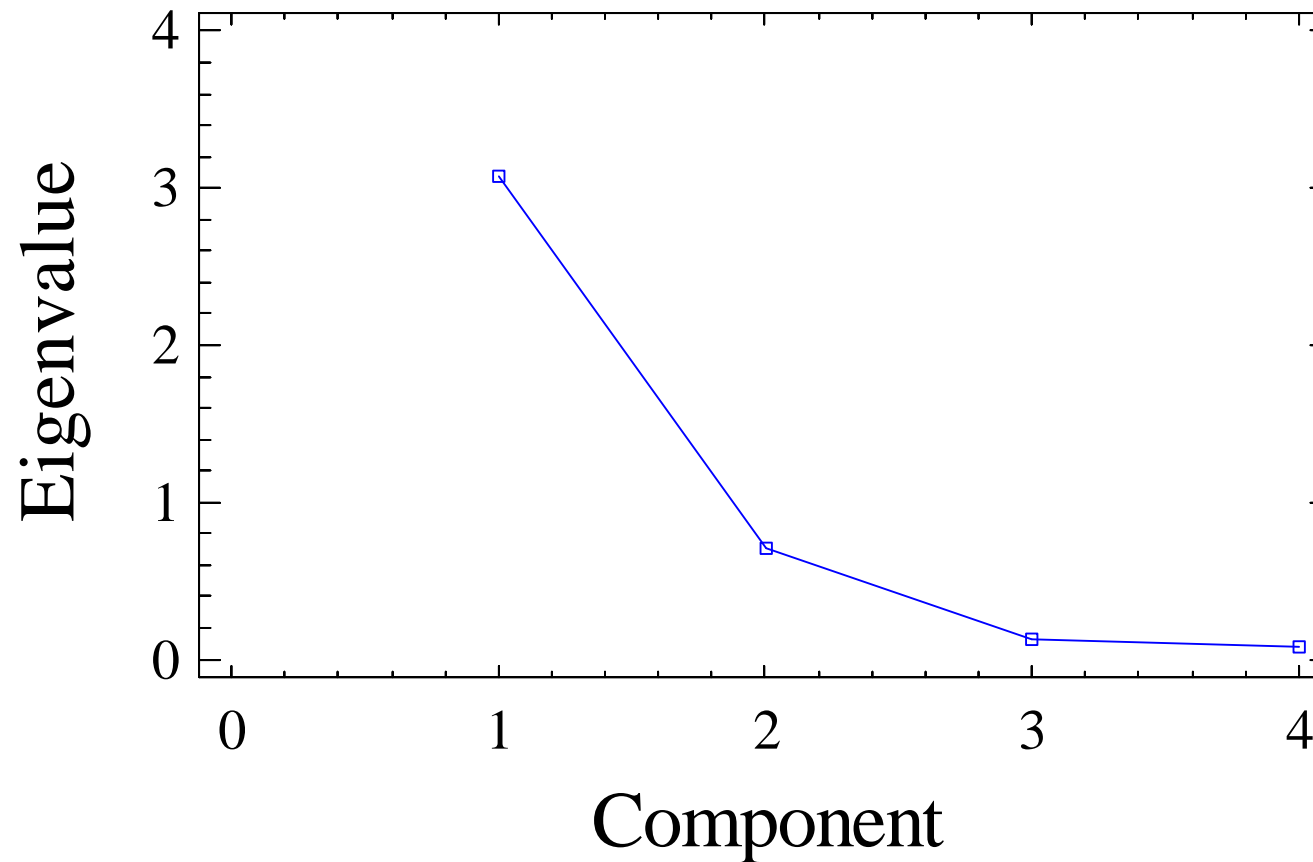
Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	3,07658	76,914	76,914
2	0,70589	17,647	94,562
3	0,13252	3,313	97,875
4	0,0850134	2,125	100,000

3. Número de componentes a retener.

- **Criterio de la media aritmética.** Se seleccionan aquellas CP cuya raíz característica (λ) supere la media de las raíces características.
 - Si tenemos variables tipificadas, todas aquellas que superen el valor 1.
- **Contraste sobre raíces no retenidas.** (suponer que las variables originales siguen una distribución normal)
- **El gráfico de sedimentación.**

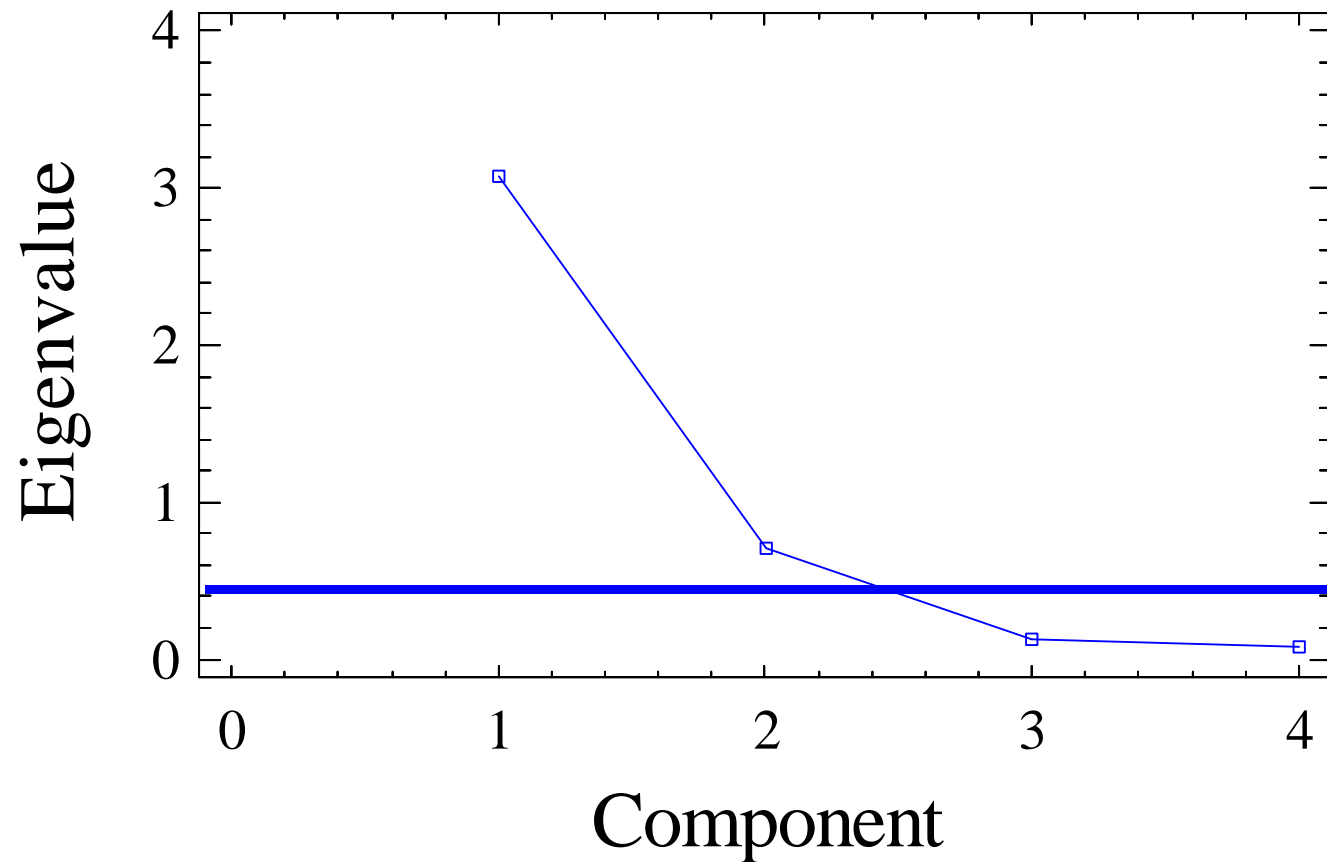
3. Número de componentes a retener.

Scree Plot



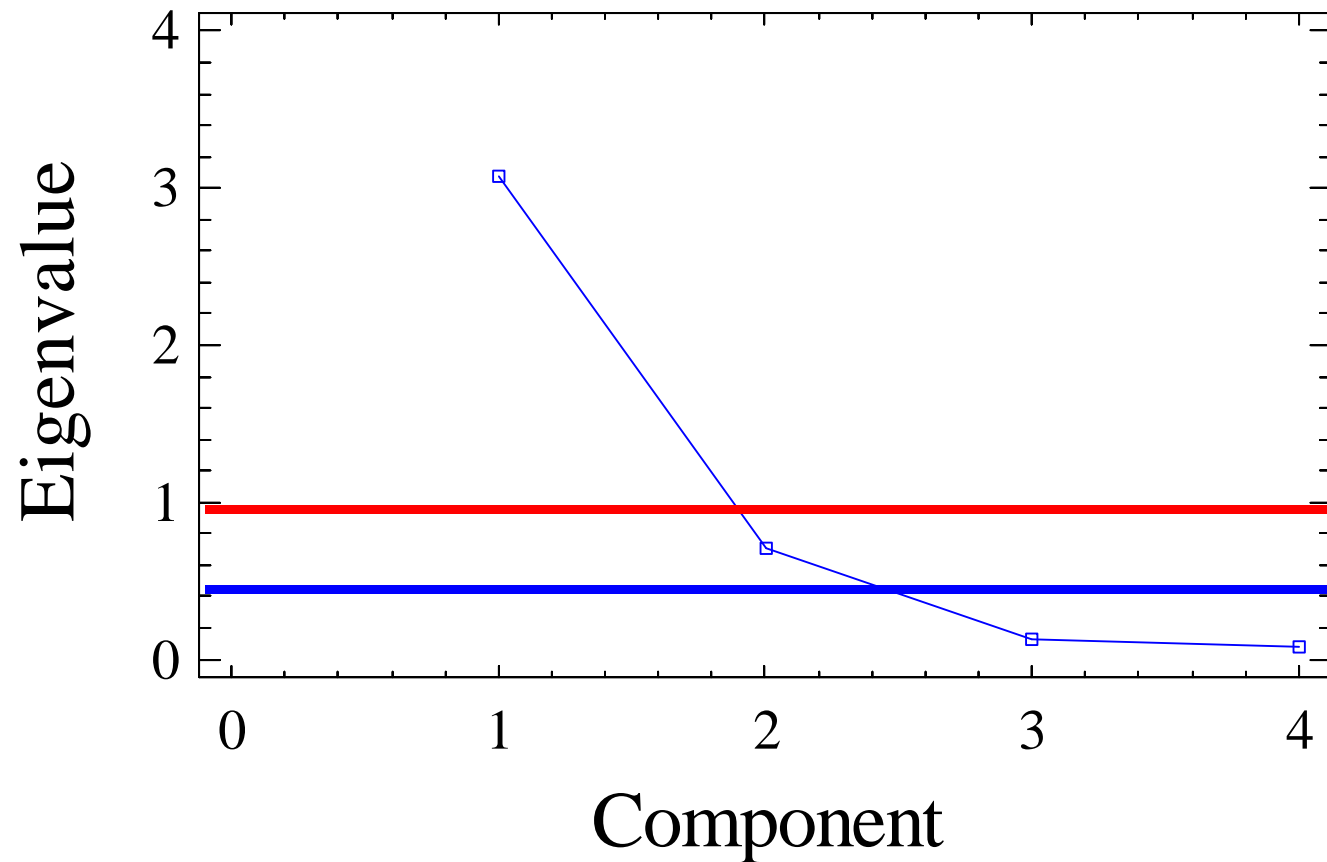
3. Número de componentes a retener.

Scree Plot



3. Número de componentes a retener.

Scree Plot



4. Retención de variables.

- Si alguna de las variables presenta correlaciones muy débiles con los componentes retenidos, lo ideal es eliminarla del estudio (no está representada en ningún CP).

4. Retención de variables.

- Si alguna de las variables presenta correlaciones muy débiles con los componentes retenidos, lo ideal es eliminarla del estudio (no está representada en ningún CP).
- Si es una variable importante a juicio del investigador, habría que retener CP hasta que dicha variable estuviese representada.

5. Interpretación de los componentes.

- El investigador debe conocer el tema en profundidad para interpretar qué parte de la variabilidad explica el CP.

5. Interpretación de los componentes.

- El investigador debe conocer el tema en profundidad para interpretar qué parte de la variabilidad explica el CP.
- Para que un CP sea fácilmente interpretable:
 - Los coeficientes deben ser próximos a 1
 - Una variable debe tener un coeficiente elevado con sólo un CP
 - No deben existir CP con coeficientes similares

Objetivo: Caracterización técnica de las explotaciones ecológicas de vacuno lechero.

Se recoge la información mediante encuestas directas al ganadero, con muestreo aleatorio estratificado según localización geográfica (12 explotaciones en total, 55% de la población).

Determinar las variables a analizar

1. Los casos deben superar al conjunto de variables

Determinar las variables a analizar

1. Los casos deben superar al conjunto de variables
2. Las variables deben:
 - Ser métricas, aunque también ficticias

Determinar las variables a analizar

1. Los casos deben superar al conjunto de variables
2. Las variables deben:
 - Ser métricas, aunque también ficticias
 - Alta variabilidad

Determinar las variables a analizar

1. Los casos deben superar al conjunto de variables
2. Las variables deben:
 - Ser métricas, aunque también ficticias
 - Alta variabilidad
 - Altas correlaciones
 - Sin dependencia lineal
 - Sin variables incorrelacionadas

Determinar las variables a analizar

1. Los casos deben superar al conjunto de variables
2. Las variables deben:
 - Ser métricas, aunque también ficticias
 - Alta variabilidad
 - Altas correlaciones
 - Sin dependencia lineal
 - Sin variables incorrelacionadas
 - Principio de parsimonia (mínimas variables)

Determinar las variables a analizar

1. Los casos deben superar al conjunto de variables
2. Las variables deben:
 - Ser métricas, aunque también ficticias
 - Alta variabilidad
 - Altas correlaciones
 - Sin dependencia lineal
 - Sin variables incorrelacionadas
 - Principio de parsimonia (mínimas variables)
 - Usar ratios en vez de absolutos

Variables originales:

- **Dimensión:** superficie (ha), número de vacas, número de empleados (UTH)
- **Estructura patrimonial:** superficie en propiedad (ha)
- **Intensificación:** consumo de concentrado (kg)
- **Productivas:** leche producida (kg) y terneros producidos
- **Gestión:** Número de terneros muertos, número de terneras de reposición, gasto sanitario

Correlaciones

	NHT	NHT_NHP	nhp	NVAC	TREP	TMORT	CARGA	ITC	ntc	ILC	NLC	ccon	CCON_NVAC	UTHT	NVAC_UTHT	GASANIT_NVAC
NHT	-	0,13	0,87	0,49	0,01	-0,11	-0,29	0,02	0,35	-0,22	0,46	0,45	-0,2	0,23	0,41	0,29
NHT_NHP	0,13	-	0,55	0,12	0,07	0,43	-0,05	-0,49	-0,02	0,72	0,32	0,09	-0,08	0,26	0,07	-0,04
nhp	0,87	0,55	-	0,54	0,12	0,07	-0,18	-0,16	0,36	0,18	0,6	0,42	-0,06	0,37	0,41	0,31
NVAC	0,49	0,12	0,54	-	0,48	-0,32	0,65	0,39	0,96	-0,02	0,96	0,52	-0,23	0,43	0,91	0,89
TREP	0,01	0,07	0,12	0,48	-	-0,47	0,62	0,45	0,58	-0,07	0,42	-0,42	-0,79	0,39	0,29	0,69
TMORT	-0,11	0,43	0,07	-0,32	-0,47	-	-0,37	-0,78	-0,47	0,53	-0,14	0,09	0,2	0,01	-0,28	-0,49
CARGA	-0,29	-0,05	-0,18	0,65	0,62	-0,37	-	0,53	0,77	0,06	0,62	0,04	-0,36	0,21	0,64	0,79
ITC	0,02	-0,49	-0,16	0,39	0,45	-0,78	0,53	-	0,59	-0,33	0,25	-0,15	-0,27	-0,03	0,37	0,66
ntc	0,35	-0,02	0,36	0,96	0,58	-0,47	0,77	0,59	-	-0,08	0,9	0,37	-0,31	0,38	0,88	0,96
ILC	-0,22	0,72	0,18	-0,02	-0,07	0,53	0,68	-0,33	-0,08	-	0,21	-0,01	0,07	-0,02	0	-0,05
NLC	0,46	0,32	0,6	0,96	0,42	-0,14	0,62	0,25	0,9	0,21	-	0,52	-0,22	0,39	0,9	0,84
ccon	0,45	0,09	0,42	0,52	-0,42	0,09	0,04	-0,15	0,37	-0,01	0,52	-	0,59	0,23	0,55	0,15
CCON_NVAC	-0,02	-0,08	-0,06	-0,23	-0,79	0,2	-0,36	-0,27	-0,31	0,07	-0,22	0,59	-	-0,01	-0,21	-0,43
UTHT	0,23	0,26	0,37	0,43	0,39	0,01	0,21	-0,03	0,38	-0,2	0,39	0,23	-0,01	-	0,07	0,34
NVAC_UTHT	0,41	0,075	0,41	0,91	0,29	-0,28	0,64	0,37	0,88	0	0,9	0,55	-0,21	0,07	-	0,79
GASANIT_NVAC	0,29	-0,44	0,31	0,89	0,69	-0,49	0,79	0,66	0,96	-0,05	0,84	0,15	-0,43	0,34	0,79	-

Correlaciones

	NHT	NHT_NHP	nhp	NVAC	TREP	TMORT	CARGA	ITC	ntc	ILC	NLC	ccon	CCON_NVAC	UTHT	NVAC_UTHT	GASANIT_NVAC
NHT	-	0,13	0,87	0,49	0,01	-0,11	-0,29	0,02	0,35	-0,22	0,46	0,45	-0,2	0,23	0,41	0,29
NHT_NHP	0,13	-	0,55	0,12	0,07	0,43	-0,05	-0,49	-0,02	0,72	0,32	0,09	-0,08	0,26	0,07	-0,04
nhp	0,87	0,55	-	0,54	0,12	0,07	-0,18	-0,16	0,36	0,18	0,6	0,42	-0,06	0,37	0,41	0,31
NVAC	0,49	0,12	0,54	-	0,48	-0,32	0,65	0,39	0,96	-0,02	0,96	0,52	-0,23	0,43	0,91	0,89
TREP	0,01	0,07	0,12	0,48	-	-0,47	0,62	0,45	0,58	-0,07	0,42	-0,42	-0,79	0,39	0,29	0,69
TMORT	-0,11	0,43	0,07	-0,32	-0,47	-	-0,37	-0,78	-0,47	0,53	-0,14	0,09	0,2	0,01	-0,28	-0,49
CARGA	-0,29	-0,05	-0,18	0,65	0,62	-0,37	-	0,53	0,77	0,06	0,62	0,04	-0,36	0,21	0,64	0,79
ITC	0,02	-0,49	-0,16	0,39	0,45	-0,78	0,53	-	0,59	-0,33	0,25	-0,15	-0,27	-0,03	0,37	0,66
ntc	0,35	-0,02	0,36	0,96	0,58	-0,47	0,77	0,59	-	-0,08	0,9	0,37	-0,31	0,38	0,88	0,96
ILC	-0,22	0,72	0,18	-0,02	-0,07	0,53	0,68	-0,33	-0,08	-	0,21	-0,01	0,07	-0,02	0	-0,05
NLC	0,46	0,32	0,6	0,96	0,42	-0,14	0,62	0,25	0,9	0,21	-	0,52	-0,22	0,39	0,9	0,84
ccon	0,45	0,09	0,42	0,52	-0,42	0,09	0,04	-0,15	0,37	-0,01	0,52	-	0,59	0,23	0,55	0,15
CCON_NVAC	-0,02	-0,08	-0,06	-0,23	-0,79	0,2	-0,36	-0,27	-0,31	0,07	-0,22	0,59	-	-0,01	-0,21	-0,43
UTHT	0,23	0,26	0,37	0,43	0,39	0,01	0,21	-0,03	0,38	-0,2	0,39	0,23	-0,01	-	0,07	0,34
NVAC_UTHT	0,41	0,075	0,41	0,91	0,29	-0,28	0,64	0,37	0,88	0	0,9	0,55	-0,21	0,07	-	0,79
GASANIT_NVAC	0,29	-0,44	0,31	0,89	0,69	-0,49	0,79	0,66	0,96	-0,05	0,84	0,15	-0,43	0,34	0,79	-

Alta correlación :

NHT con nhp

NVAC con ntc

NVAC con NLC

NVAC_UTH con NVAC

NVAC con GASANIT_VAC

NLC con ntc

ntc con GASANIT_VAC

Ninguna incorrelacionada o con coeficientes muy bajos

Descripción estadística

Summary Statistics

	NHT	NHT_NHP	NVAC
Count	12	12	12
Average	42,8108	0,592407	52,8333
Median	37,5	0,54	42,0
Variance	1003,82	0,110589	1322,52
Standard error	9,14615	0,0959989	10,4981
Minimum	18,0	0,0555556	14,0
Maximum	136,0	1,0	128,0
Lower quartile	24,5	0,341667	29,0
Upper quartile	48,0	0,899451	62,5
Coeff. of variation	74,0074%	56,1354%	68,8323%

	NREP	NTM	TREP
Count	12	12	12
Average	25,0833	3,25	20,1842
Median	15,5	3,0	22,16
Variance	1020,08	4,56818	122,09
Standard error	9,21992	0,616994	3,18969
Minimum	1,0	0,0	0,76
Maximum	109,0	6,0	42,58
Lower quartile	6,5	1,5	14,105
Upper quartile	21,0	5,5	24,29
Coeff. of variation	127,331%	65,764%	54,743%

Descripción estadística

	NTP	TMORT	CARGA
Count	12	12	12
Average	48,5833	8,28083	1,57167
Median	35,0	8,535	1,22
Variance	1636,81	23,2112	2,54731
Standard error	11,6791	1,39078	0,460734
Minimum	14,0	0,0	0,37
Maximum	148,0	15,15	6,49
Lower quartile	25,0	5,335	0,945
Upper quartile	51,0	11,655	1,47
Coeff. of variation	83,2745%	58,1801%	101,55%

	ITC	ILC	NLC
Count	12	12	12
Average	0,820833	5351,34	281543,0
Median	0,86	5704,68	228850,0
Variance	0,0234811	2,15118E6	3,8528E10
Standard error	0,0442352	423,397	56662,7
Minimum	0,57	2500,0	50000,0
Maximum	1,14	7571,43	650000,0
Lower quartile	0,69	4836,54	152200,0
Upper quartile	0,89	6325,4	353209,0
Coeff. of variation	18,6683%	27,4079%	69,7177%

Descripción estadística

	CCON_NVAC	CCON_NVAC*NVAC	UTHT
Count	12	12	12
Average	1899,78	92601,7	2,28
Median	1525,0	65575,0	2,0
Variance	1,01814E6	4,52938E9	0,6222
Standard error	291,282	19428,0	0,227706
Minimum	698,48	21000,0	1,16
Maximum	3406,62	221430,0	4,0
Lower quartile	1143,75	47427,4	2,0
Upper quartile	2998,81	137000,0	2,895
Coeff. of variation	53,113%	72,6776%	34,5963%

	GASANIT_NVAC	GASANIT_NVAC*NVAC
Count	12	12
Average	6558,33	719057,0
Median	1650,0	49500,0
Variance	1,56612E8	2,62893E12
Standard error	3612,62	468057,0
Minimum	0,0	0,0
Maximum	40200,0	5,1456E6
Lower quartile	1000,0	26900,0
Upper quartile	2640,0	144240,0
Coeff. of variation	190,818%	225,49%

Descripción estadística

Summary Statistics

	NHT	NHT_NHP	NVAC
Count	12	12	12
Average	42,8108	0,592407	52,8333
Median	37,5	0,54	42,0
Variance	1003,82	0,110589	1322,52
Standard error	9,14615	0,0959989	10,4981
Minimum	18,0	0,0555556	14,0
Maximum	136,0	1,0	128,0
Lower quartile	24,5	0,341667	29,0
Upper quartile	48,0	0,899451	62,5
Coeff. of variation	74,0074%	56,1354%	68,8323%

	NREP	NTM	TREP
Count	12	12	12
Average	25,0833	3,25	20,1842
Median	15,5	3,0	22,16
Variance	1020,08	4,56818	122,09
Standard error	9,21992	0,616994	3,18969
Minimum	1,0	0,0	0,76
Maximum	109,0	6,0	42,58
Lower quartile	6,5	1,5	14,105
Upper quartile	21,0	5,5	24,29
Coeff. of variation	127,331%	65,764%	54,743%

Descripción estadística

	NTP	TMORT	CARGA
Count	12	12	12
Average	48,5833	8,28083	1,57167
Median	35,0	8,535	1,22
Variance	1636,81	23,2112	2,54731
Standard error	11,6791	1,39078	0,460734
Minimum	14,0	0,0	0,37
Maximum	148,0	15,15	6,49
Lower quartile	25,0	5,335	0,945
Upper quartile	51,0	11,655	1,47
Coeff. of variation	83,2745%	58,1801%	101,55%

	ITC	ILC	NLC
Count	12	12	12
Average	0,820833	5351,34	281543,0
Median	0,86	5704,68	228850,0
Variance	0,0234811	2,15118E6	3,8528E10
Standard error	0,0442352	423,397	56662,7
Minimum	0,57	2500,0	50000,0
Maximum	1,14	7571,43	650000,0
Lower quartile	0,69	4836,54	152200,0
Upper quartile	0,89	6325,4	353209,0
Coeff. of variation	18,6683%	27,4079%	69,7177%

Descripción estadística

	CCON_NVAC	CCON_NVAC*NVAC	UTHT
Count	12	12	12
Average	1899,78	92601,7	2,28
Median	1525,0	65575,0	2,0
Variance	1,01814E6	4,52938E9	0,6222
Standard error	291,282	19428,0	0,227706
Minimum	698,48	21000,0	1,16
Maximum	3406,62	221430,0	4,0
Lower quartile	1143,75	47427,4	2,0
Upper quartile	2998,81	137000,0	2,895
Coeff. of variation	53,113%	72,6776%	34,5963%

	GASANIT_NVAC	GASANIT_NVAC*NVAC
Count	12	12
Average	6558,33	719057,0
Median	1650,0	49500,0
Variance	1,56612E8	2,62893E12
Standard error	3612,62	468057,0
Minimum	0,0	0,0
Maximum	40200,0	5,1456E6
Lower quartile	1000,0	26900,0
Upper quartile	2640,0	144240,0
Coeff. of variation	190,818%	225,49%

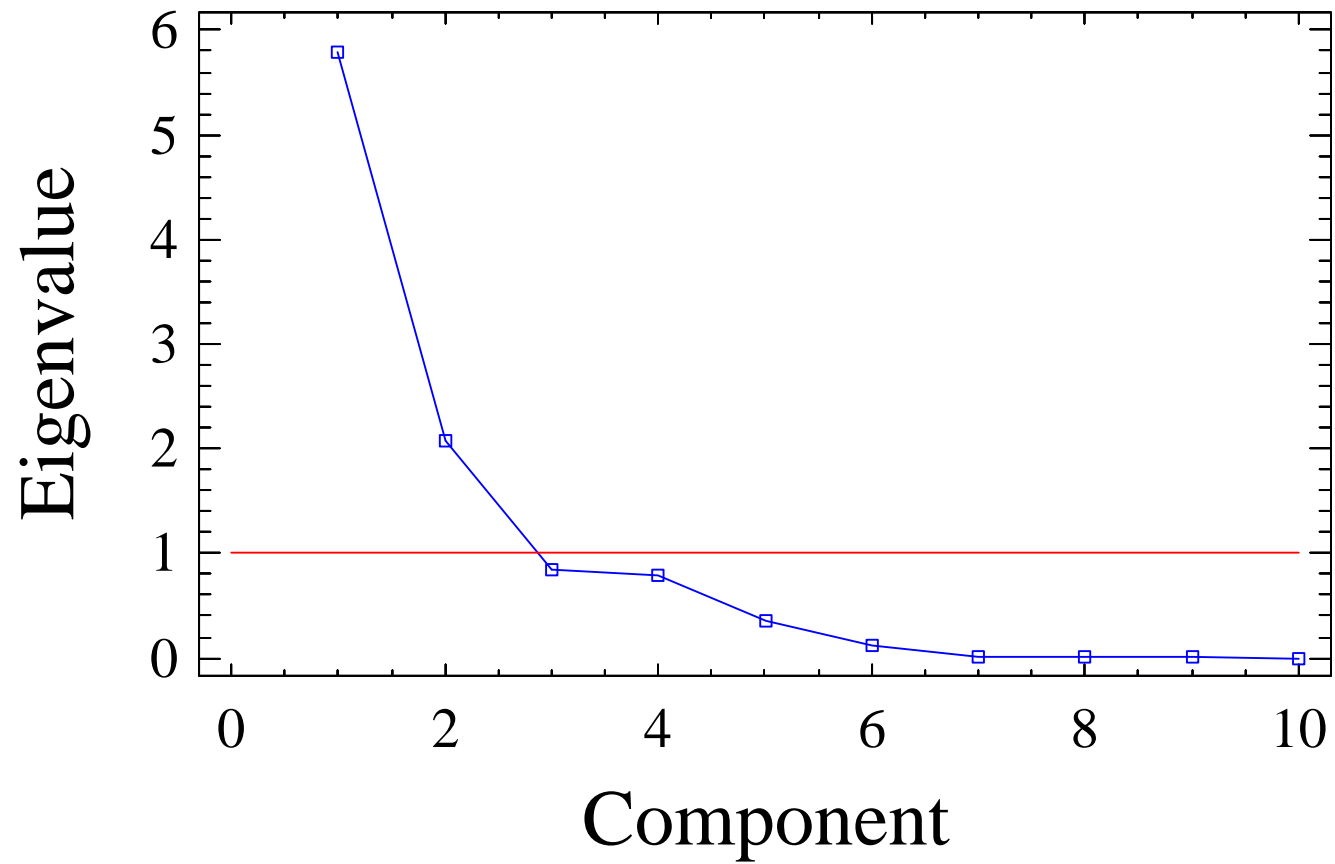
Variables seleccionadas:

- **Dimensión:** superficie (ha), número de vacas, número de empleados (UTH)
- **Estructura patrimonial:** superficie en propiedad (ha)
- **Intensificación:** consumo de concentrado (kg)
- **Productivas:** leche producida (kg) y terneros producidos
- **Gestión:** Número de terneros muertos, número de terneras de reposición, gasto sanitario

Principal Components Analysis

Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	5,79804	57,980	57,980
2	2,07674	20,767	78,748
3	0,844575	8,446	87,194
4	0,786216	7,862	95,056
5	0,356381	3,564	98,620
6	0,112722	1,127	99,747
7	0,0159819	0,160	99,907
8	0,00609229	0,061	99,967
9	0,00301224	0,030	99,998
10	0,000239632	0,002	100,000

Scree Plot



Matriz factorial

	CP1	CP2
NHT	0,61	-0,54
NHP	0,68	-0,53
NVAC	0,97	0,13
UTH	0,51	-0,03
CCON	0,54	0,55
GASAN	0,84	0,51
NTP	0,92	0,31
NLC	0,96	0,05
NREP	0,82	0,54
NTM	0,49	-0,7

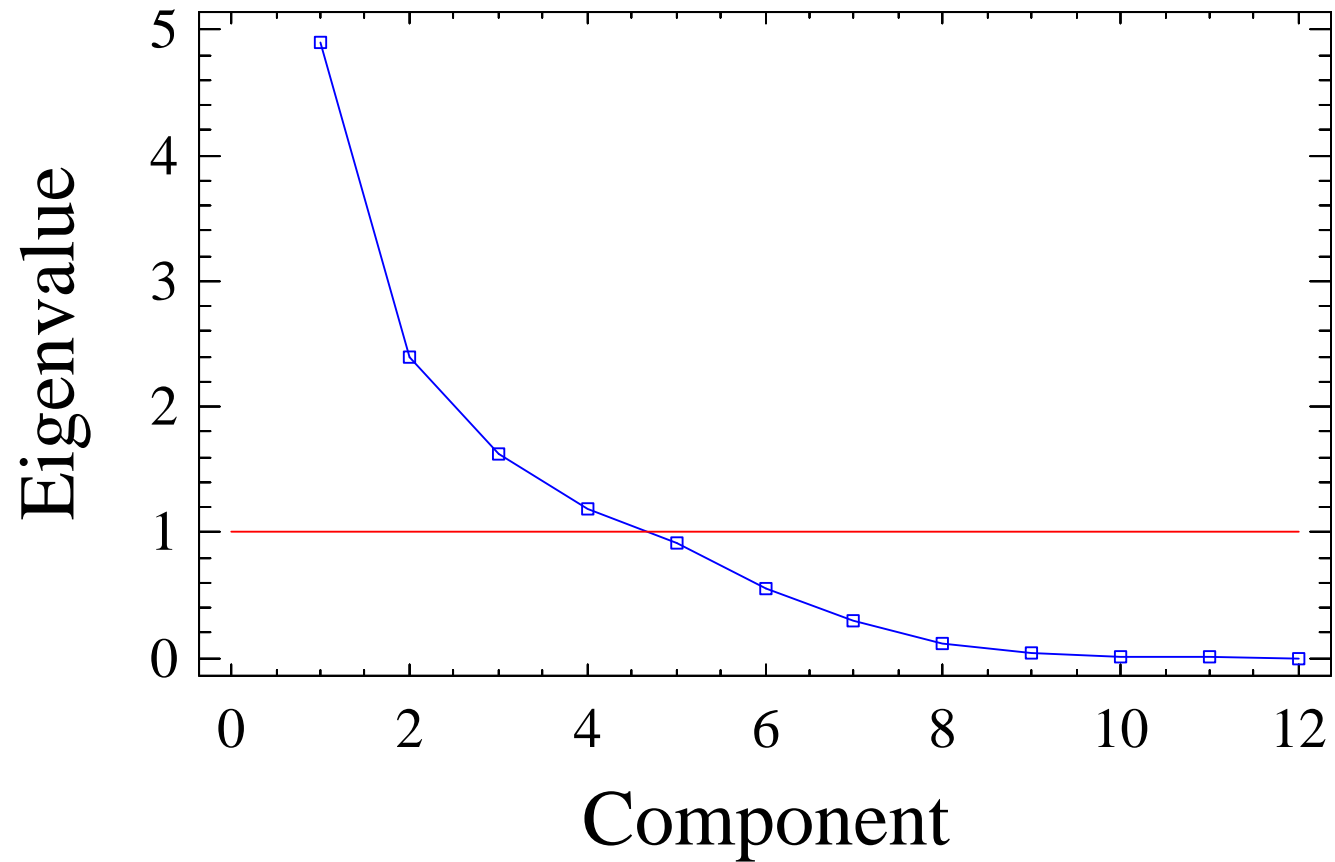
Variables seleccionadas:

- **Dimensión:** superficie (ha), número de vacas, número de empleados (UTH)
- **Estructura patrimonial:** superficie en propiedad (%)
- **Intensificación:** consumo de concentrado **por vaca** (kg), carga ganadera, **UTH/número de vacas**
- **Productivas:** leche producida **por vaca** (kg) y terneros producidos **por vaca**
- **Gestión:** Tasa de mortalidad, Tasa de reposición, Gasto sanitario **por vaca**

Principal Components Analysis

Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	4,89584	40,799	40,799
2	2,3935	19,946	60,744
3	1,61807	13,484	74,228
4	1,18132	9,844	84,073
5	0,916237	7,635	91,708
6	0,546603	4,555	96,263
7	0,292176	2,435	98,698
8	0,109103	0,909	99,607
9	0,037093	0,309	99,916
10	0,00903205	0,075	99,991
11	0,00102991	0,009	100,000
12	0,0	0,000	100,000

Scree Plot



Matriz factorial

	CP1	CP2	CP3	CP4	CP5
NHT	0,25	0,18	-0,83	-0,28	-0,31
NHT/NHP	-0,12	0,89	0,08	-0,1	-0,12
NVAC	0,85	0,35	-0,33	0,14	0,05
UTHT	0,31	0,4	-0,18	-0,47	0,67
TREP	0,76	0,09	0,41	-0,44	-0,2
TMORT	-0,64	0,58	0,04	0,05	0
ITC	0,72	-0,5	0,08	0,15	0
CARGA	0,79	0,1	0,39	0,33	0,23
ILC	-0,21	0,75	0,35	0,33	-0,07
CCON/NVAC	-0,52	-0,06	-0,46	0,46	0,48
NVAC/UTH	0,76	0,27	-0,3	0,4	-0,19
GASAN/NVAC	0,96	0,14	-0,03	0,09	0,02

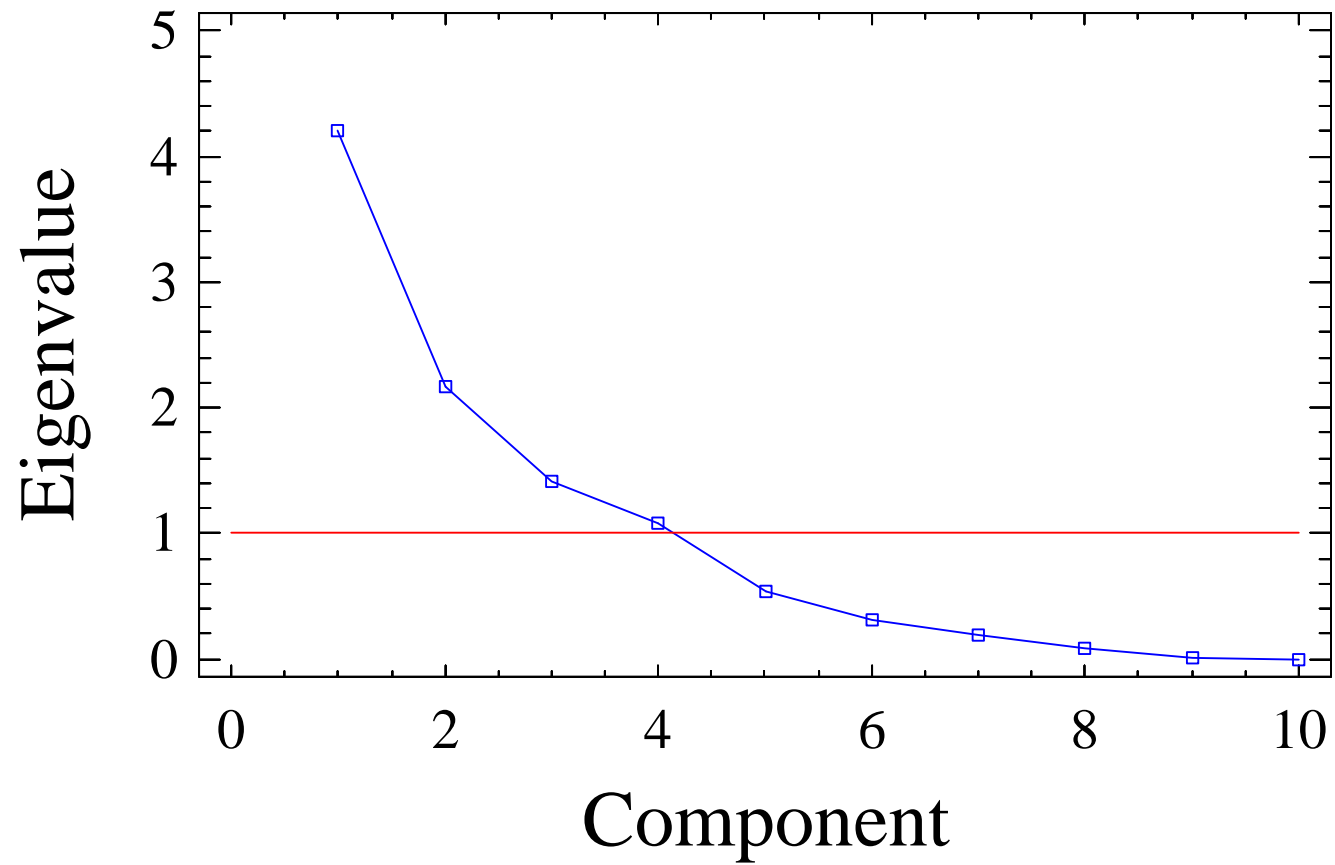
Variables seleccionadas:

- **Dimensión:** superficie (ha), ~~número de vacas~~, ~~número de empleados (UTH)~~
- **Estructura patrimonial:** superficie en propiedad (%)
- **Intensificación:** consumo de concentrado **por vaca** (kg), carga ganadera, **UTH/número de vacas**
- **Productivas:** leche producida **por vaca** (kg) y terneros producidos **por vaca**
- **Gestión:** Tasa de mortalidad, Tasa de reposición, Gasto sanitario **por vaca**

Principal Components Analysis

Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	4,20154	42,015	42,015
2	2,1611	21,611	63,626
3	1,41165	14,116	77,743
4	1,08163	10,816	88,559
5	0,535232	5,352	93,912
6	0,311971	3,120	97,031
7	0,19467	1,947	98,978
8	0,0899084	0,899	99,877
9	0,0118104	0,118	99,995
10	0,000486311	0,005	100,000

Scree Plot



Matriz factorial

	CP1	CP2	CP3	CP4	CP5
NHT	-0,14	0	0,94	-0,27	0,06
NHT/NHP	0,26	0,85	0,15	-0,12	0,27
TREP	-0,78	0,29	-0,24	-0,42	0,06
TMORT	0,73	0,46	0	0,02	-0,4
ITC	-0,81	-0,33	-0,08	0,16	0,28
CARGA	-0,78	0,3	-0,3	0,38	-0,17
ILC	0,3	0,81	-0,15	0,25	0,27
CCON/NVAC	0,57	-0,3	0,22	0,64	0,19
NVAC/UTH	-0,66	0,31	0,47	0,35	-0,24
GASAN/NVAC	-0,91	0,25	0,19	0,15	-0,02

Matriz factorial

	CP1	CP2	CP3	CP4	CP5
NHT	-0,14	0	0,94	-0,27	0,06
NHT/NHP	0,26	0,85	0,15	-0,12	0,27
TREP	-0,78	0,29	-0,24	-0,42	0,06
TMORT	0,73	0,46	0	0,02	-0,4
ITC	-0,81	-0,33	-0,08	0,16	0,28
CARGA	-0,78	0,3	-0,3	0,38	-0,17
ILC	0,3	0,81	-0,15	0,25	0,27
CCON/NVAC	0,57	-0,3	0,22	0,64	0,19
NVAC/UTH	-0,66	0,31	0,47	0,35	-0,24
GASAN/NVAC	-0,91	0,25	0,19	0,15	-0,02

Caso práctico.

1. Caracterizar y tipificar a los alumnos del master.
 - Aplicar CP
 - Interpretar
 - Aplicar Cluster
 - Interpretar