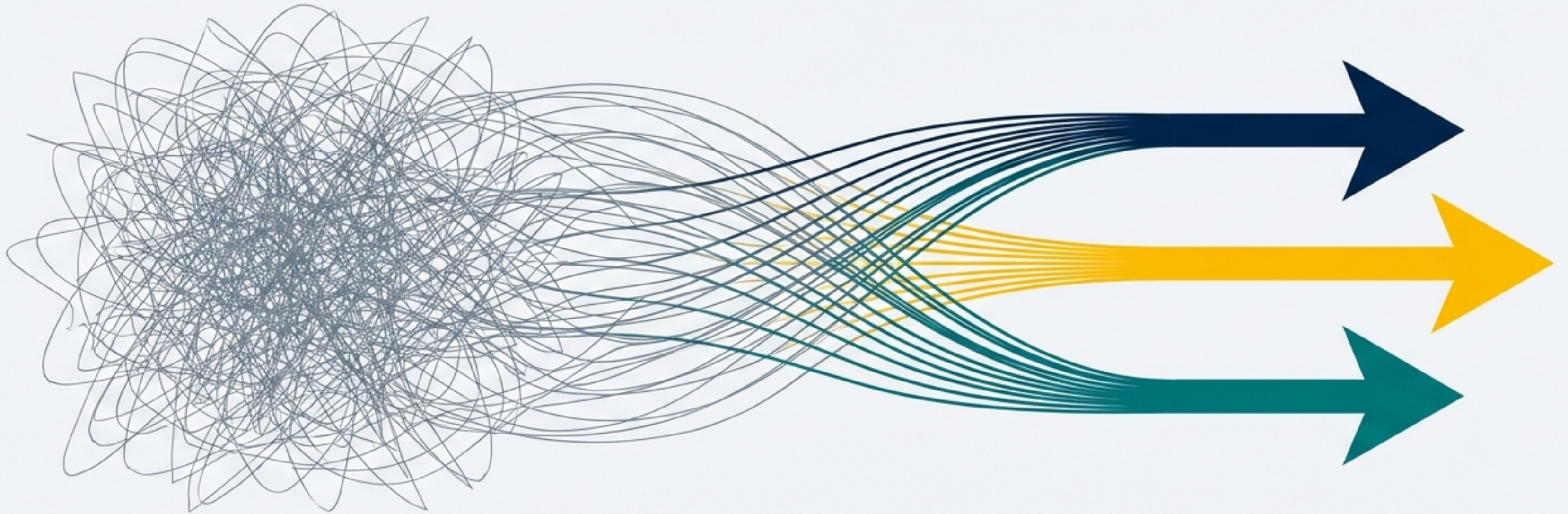


Análisis de Componentes Principales (ACP)

De la complejidad a la claridad: revelando la estructura oculta de los datos.



Una guía técnica y visual para la reducción de dimensionalidad y la toma de decisiones estratégicas.

El Desafío: Cuando el exceso de información se convierte en ruido.

Variables Originales (Ruido)



El Problema

Trabajar con grandes volúmenes de datos a menudo implica información redundante. Las variables están correlacionadas, midiendo indirectamente lo mismo.

Factores (Señal)



La Solución (ACP)

Una técnica de análisis multivariante que transforma un conjunto de variables correlacionadas en un número menor de variables no correlacionadas (factores), simplificando la realidad sin perder información crítica.

Exceso de variables → Redundancia → Parálisis en el análisis

La Lista de Verificación: ¿Son sus datos aptos para el ACP?

01

✓ Tipo de Variable

Ideal: Variables Métricas (cuantitativas).

Precaución: Las variables categóricas requieren tratamientos especiales.

02

✓ Tamaño de la Muestra

Mínimo absoluto: 50 observaciones.

Recomendado: >100 observaciones.

Ratio de Oro: 10 observaciones por cada variable analizada.

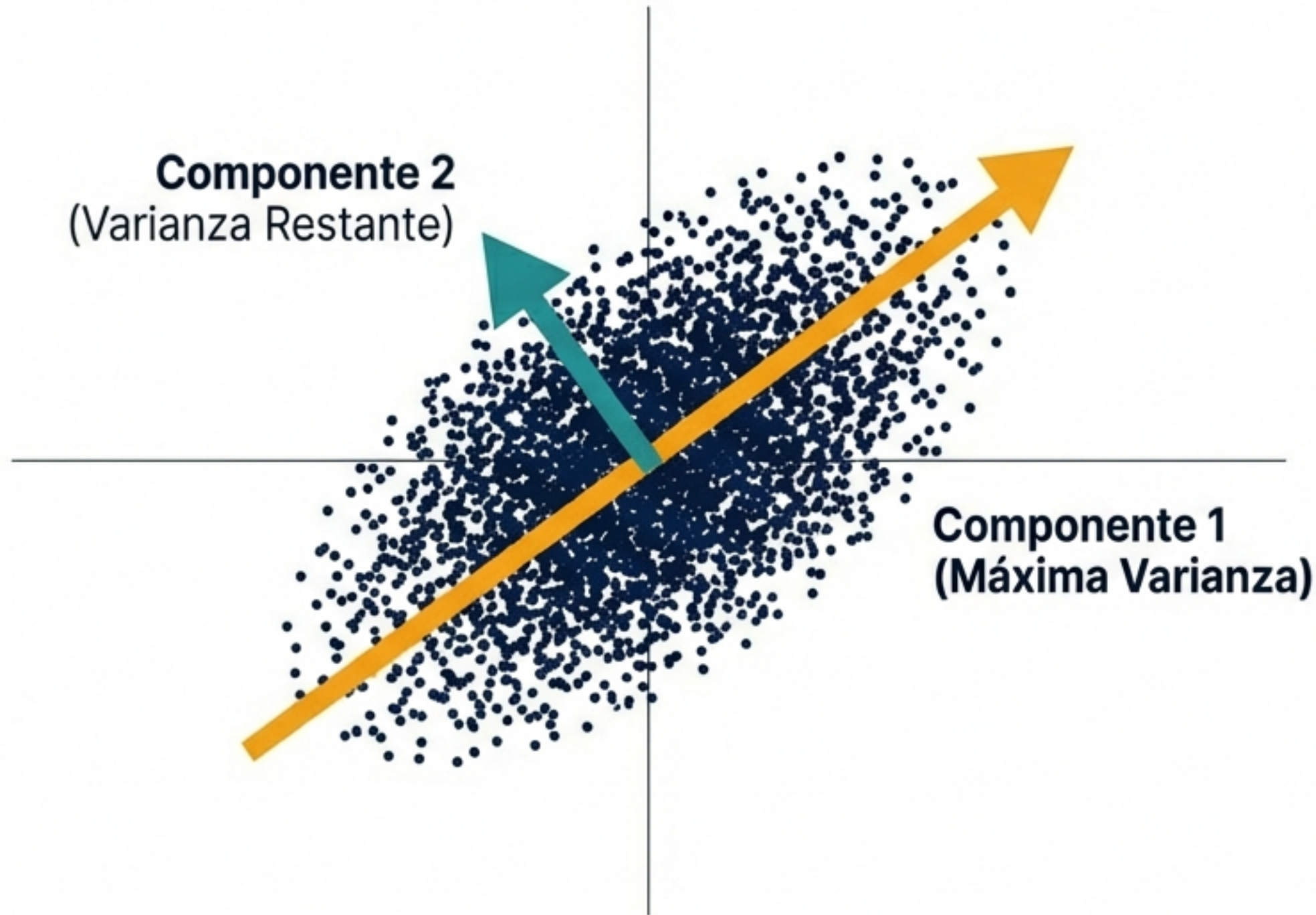
03

✓ La Prueba de Fuego

Requisito: Correlaciones significativas (> 0.3).

Insight: Si las variables no están relacionadas (incorrelación), no hay estructura que reducir. Obtendríamos tantos factores como variables.

La Mecánica: Capturando la máxima variabilidad posible.

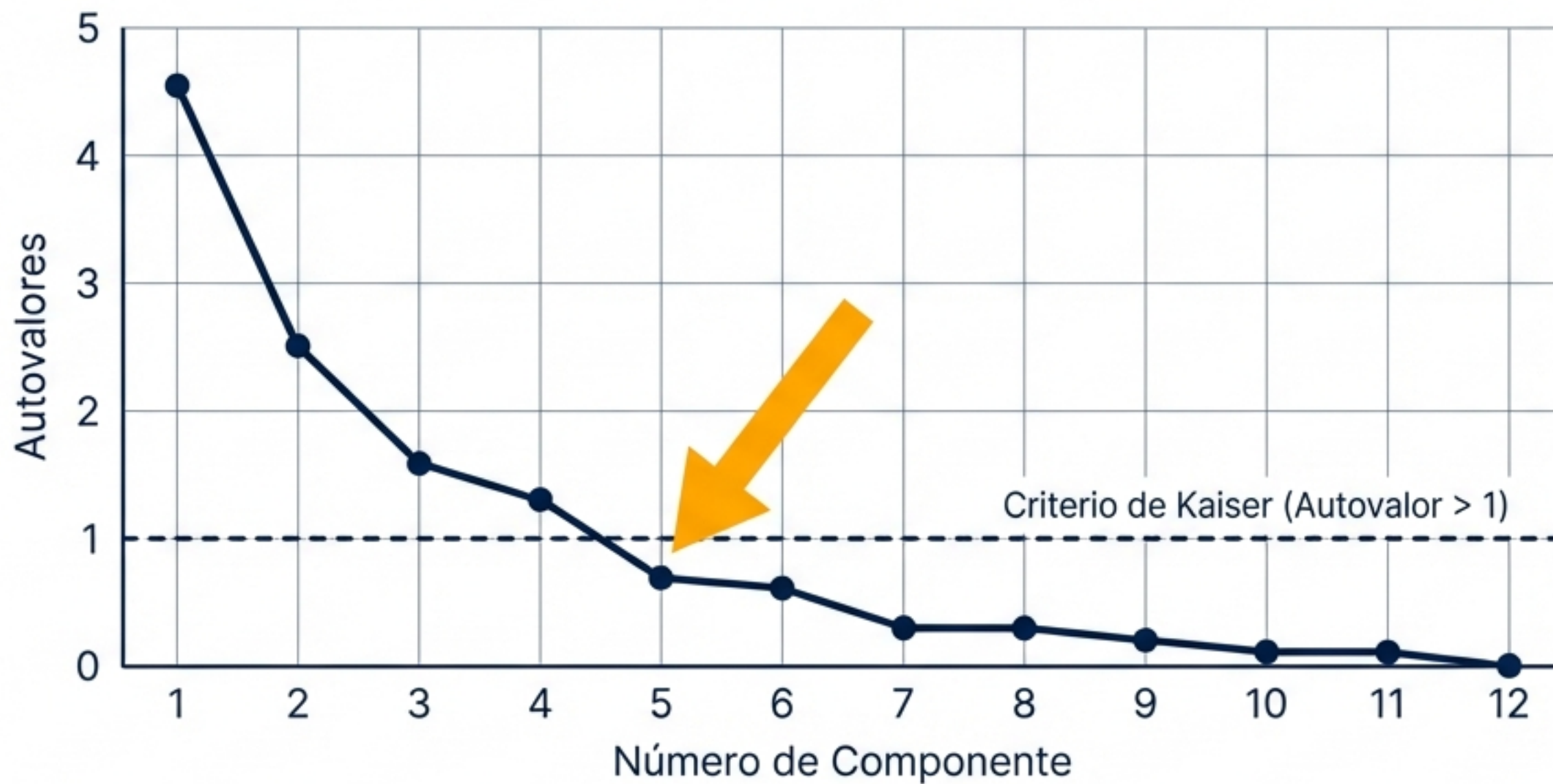


Vectores y Valores Propios: El motor matemático del ACP.

- **Componente 1:** Se elige para explicar la mayor proporción posible de la variabilidad original de los datos.
- **Componente 2 y sucesivos:** Cada nuevo componente explica lo que el anterior dejó fuera, garantizando que sean independientes entre sí (ortogonales).

El Punto de Corte: ¿Cuántos factores debemos conservar?

Gráfico de sedimentación

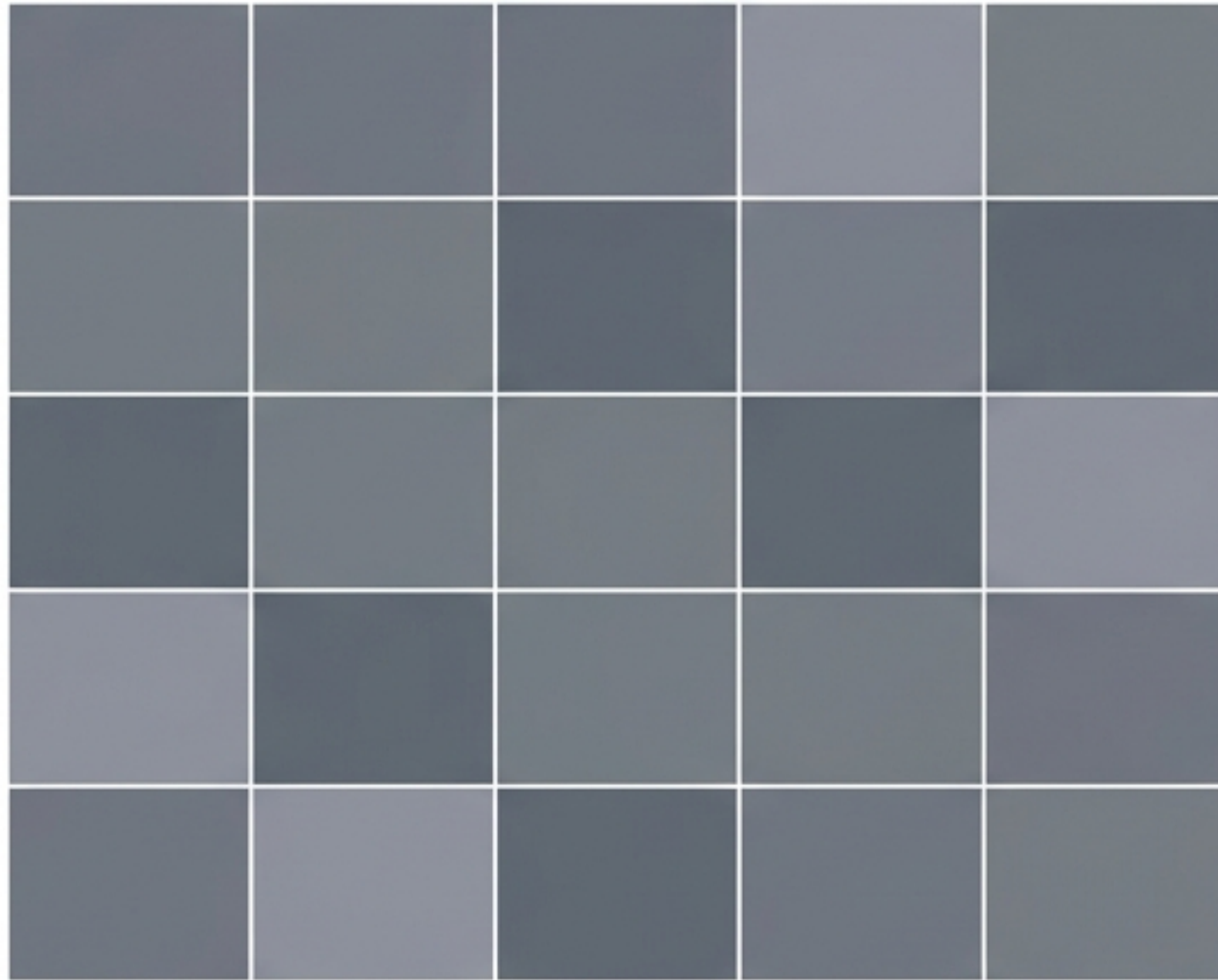


Reglas de Decisión

- 1 **Criterio de Kaiser:** Conservar solo componentes con Autovalores > 1 .
- 2 **Regla del Codo:** Identificar el punto de inflexión en el gráfico de sedimentación.
- 3 **Varianza Acumulada:** Buscar explicar entre el 60% y el 80% de la varianza total.

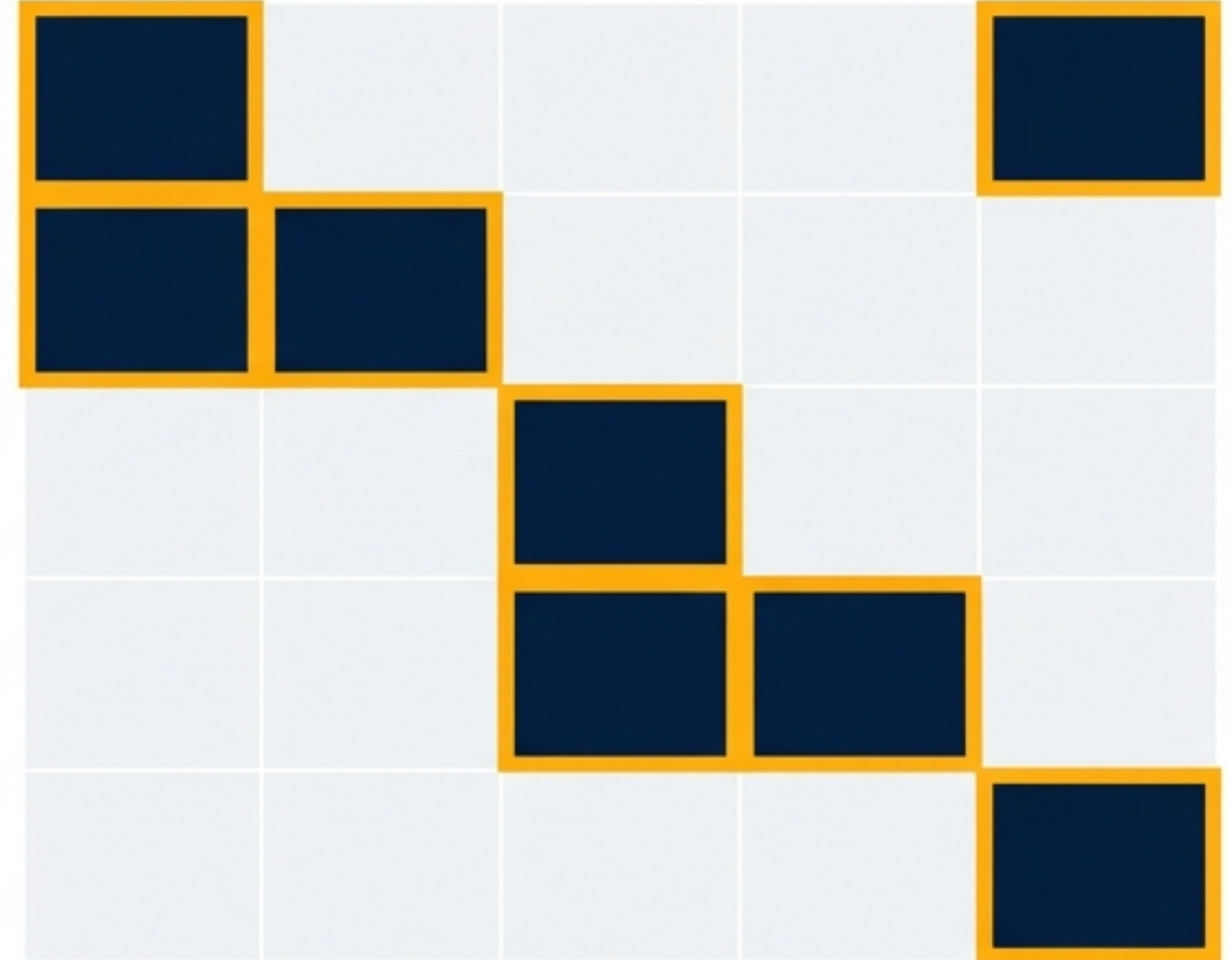
La Magia de la Rotación: Enfocando la imagen.

Matriz Sin Rotar



Cargas ambiguas y difusas.

Rotación Varimax



Estructura Simple y Polarizada.

El Problema: La matriz inicial suele ser ambigua.

La Solución (Varimax): Rotamos los ejes para polarizar las cargas.

El Objetivo: Estructura Simple. Cada variable correlaciona alto con un solo factor.

Lectura Técnica: Caso de Ingeniería de Puentes.

	Variable	Factor 1	Factor 2	
1	Luz Principal	.953		Signal Amber Dimensiones Estructurales Inter
2	Canto Tablero	.937		
3	Armadura Activa	.919	.216	
4	Número de Vanos	-.267	-.861	Deep Navy Configuración Geométrica Inter
5	Apoyos Estribo		.836	

La “Carga Factorial” (ej. 0.953) es la correlación entre la variable y el factor. Es el peso de esa variable en el concepto.

Interpretación Semántica: Dando nombre a lo invisible.

FACTOR 1

Estabilidad Financiera

- + Edad
- + Empleo
- + Ahorros
- + Residencia

Lógica: Explica el largo plazo y la acumulación de recursos.

FACTOR 2

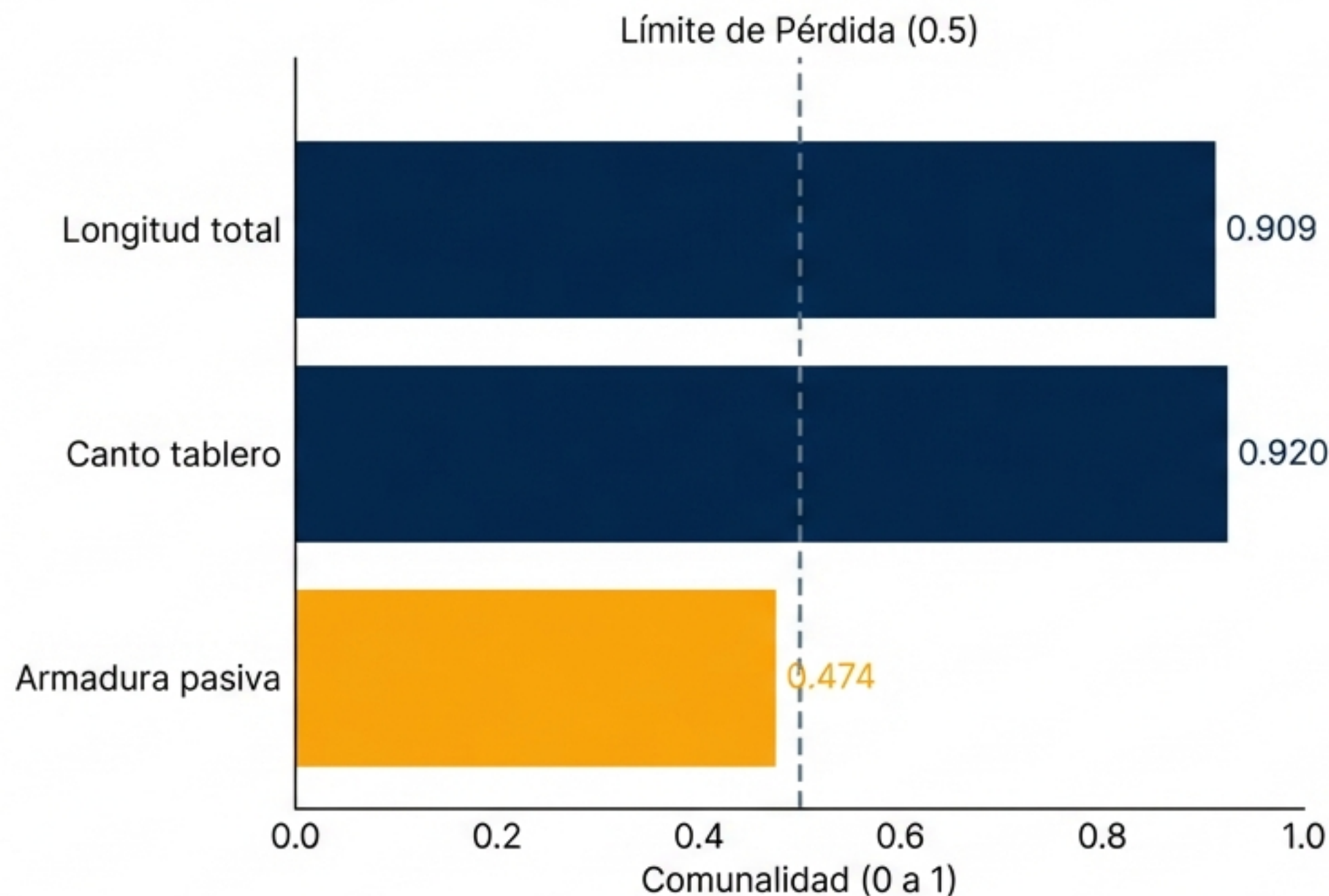
Historial Crediticio

- Deuda
- Tarjetas de Crédito

Lógica: Explica el comportamiento de pasivos y riesgo.

El algoritmo agrupa los números; el analista otorga el significado de negocio.

Control de Calidad: Las Comunalidades.



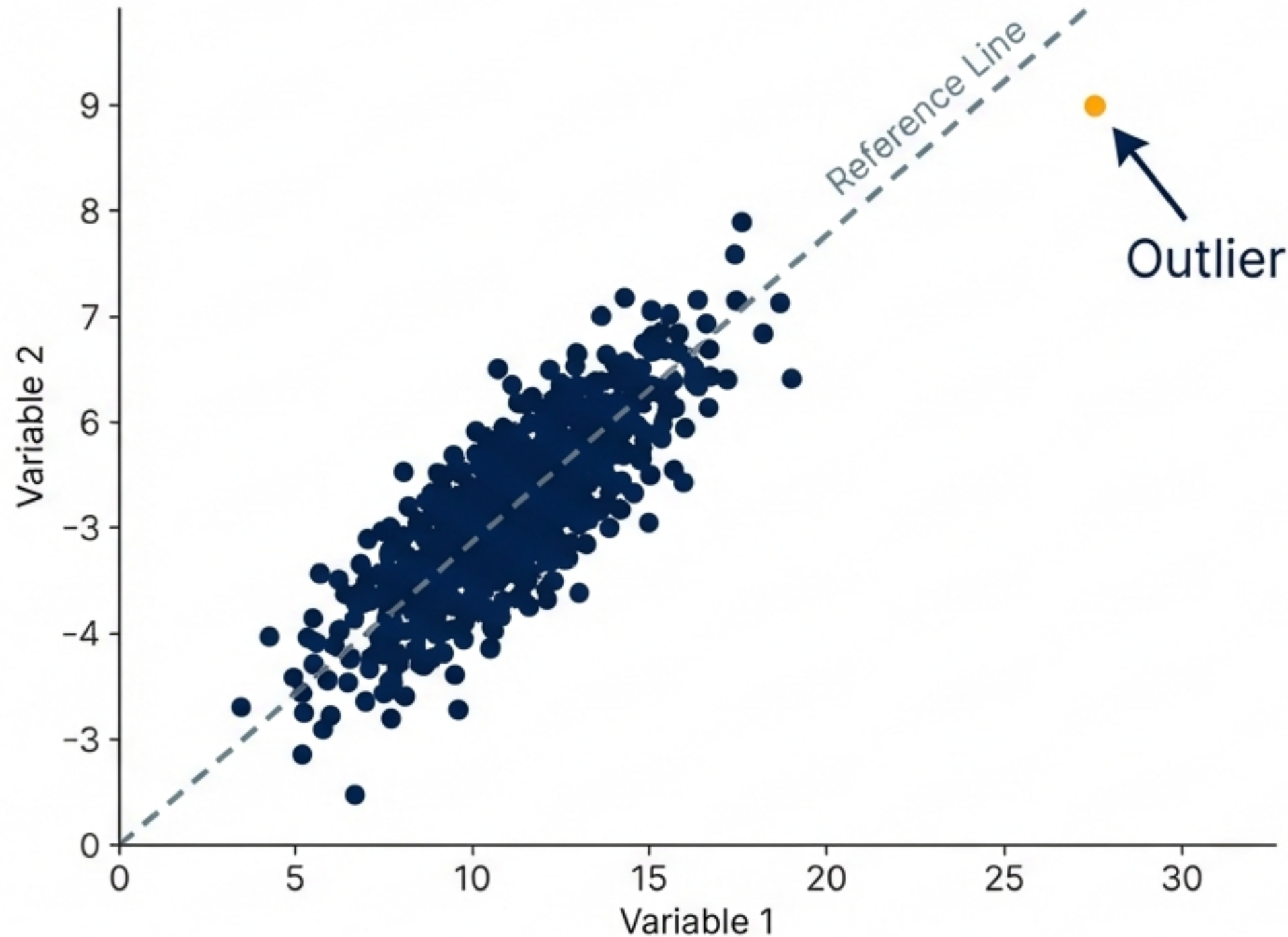
- **Definición:** La comunalidad mide qué proporción de la varianza de una variable es explicada por el modelo.
- **La Regla:** Valores < 0.5 indican que la variable se 'pierde' en la traducción.
- **Acción:** Si la comunalidad es baja (como Armadura pasiva), considere eliminar la variable.

El Contexto Importa: Tamaño de Muestra vs. vs. Carga Significativa



A menor cantidad de datos, mayor debe ser la evidencia estadística (correlación) para considerar una variable como parte de un factor.

Detección de Anomalías: Encontrando lo atípico.



Uso Alternativo: El ACP no solo resume, también expone lo que no encaja.

Valores Atípicos (Outliers):
Puntos fuera de la estructura de correlación normal.

Protocolo:

1. Revisar error de datos.
2. Analizar si es un caso especial.
3. Considerar eliminar la observación.

Escalamiento: ACP en Big Data (Google BigQuery)

```
CREATE MODEL dataset.nombre_modelo
OPTIONS(
  model_type='PCA',
  num_principal_components=3,
  scale_features=TRUE
) AS
SELECT * EXCEPT(is_male) FROM dataset.natality
```



El Flujo de Trabajo del Analista: Resumen.



De los Datos a la Estrategia.



El Análisis de Componentes Principales nos permite ver el bosque sin perdernos en los árboles.

Evalúe sus datos. Reduzca el ruido. Encuentre la señal.