



Objetivo

Análisis factorial

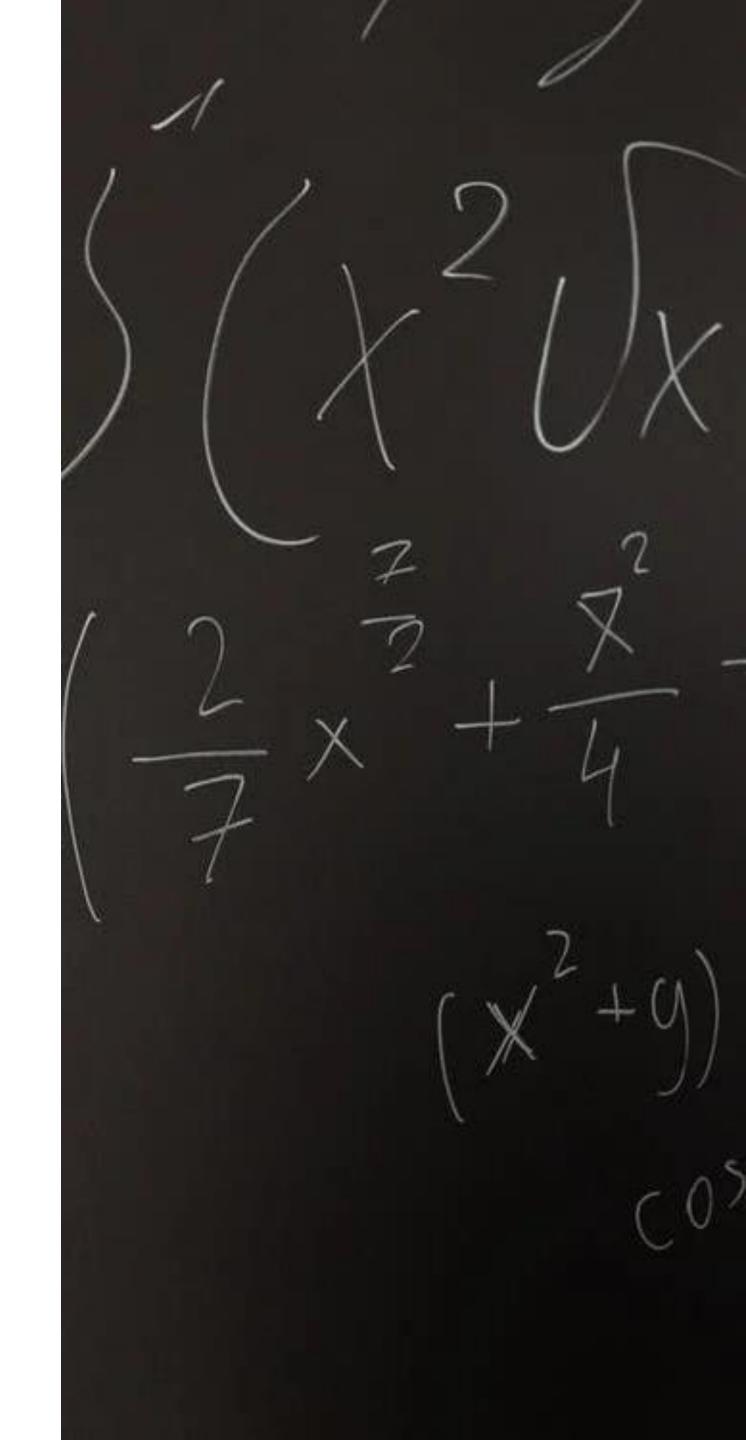
- Comprender el análisis factorial.
- Diferenciar el análisis factorial de otras técnicas de análisis multivariante
- Resolver un problema utilizando python y un conjunto de datos de UCI Machine Learning Repository.



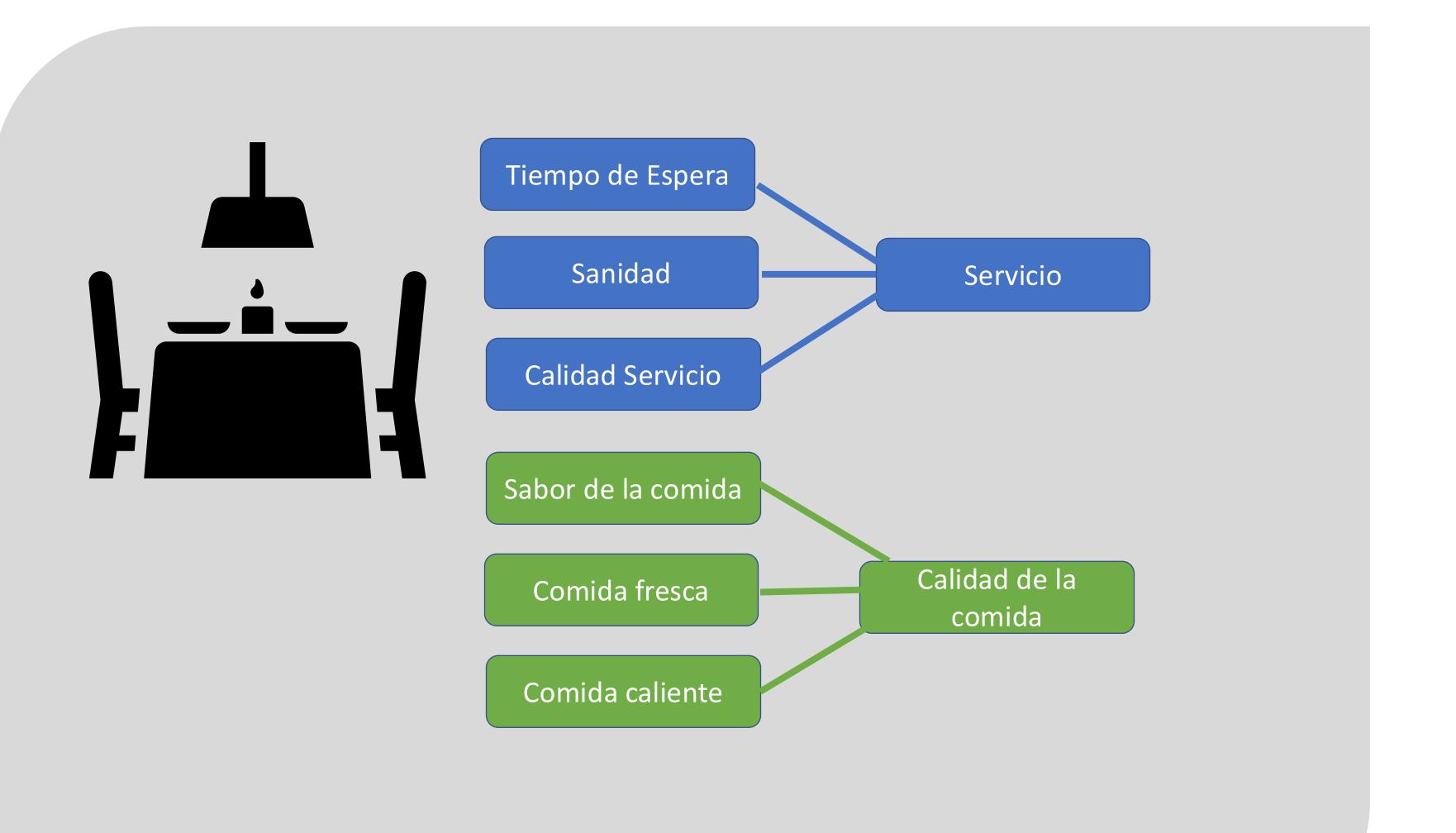


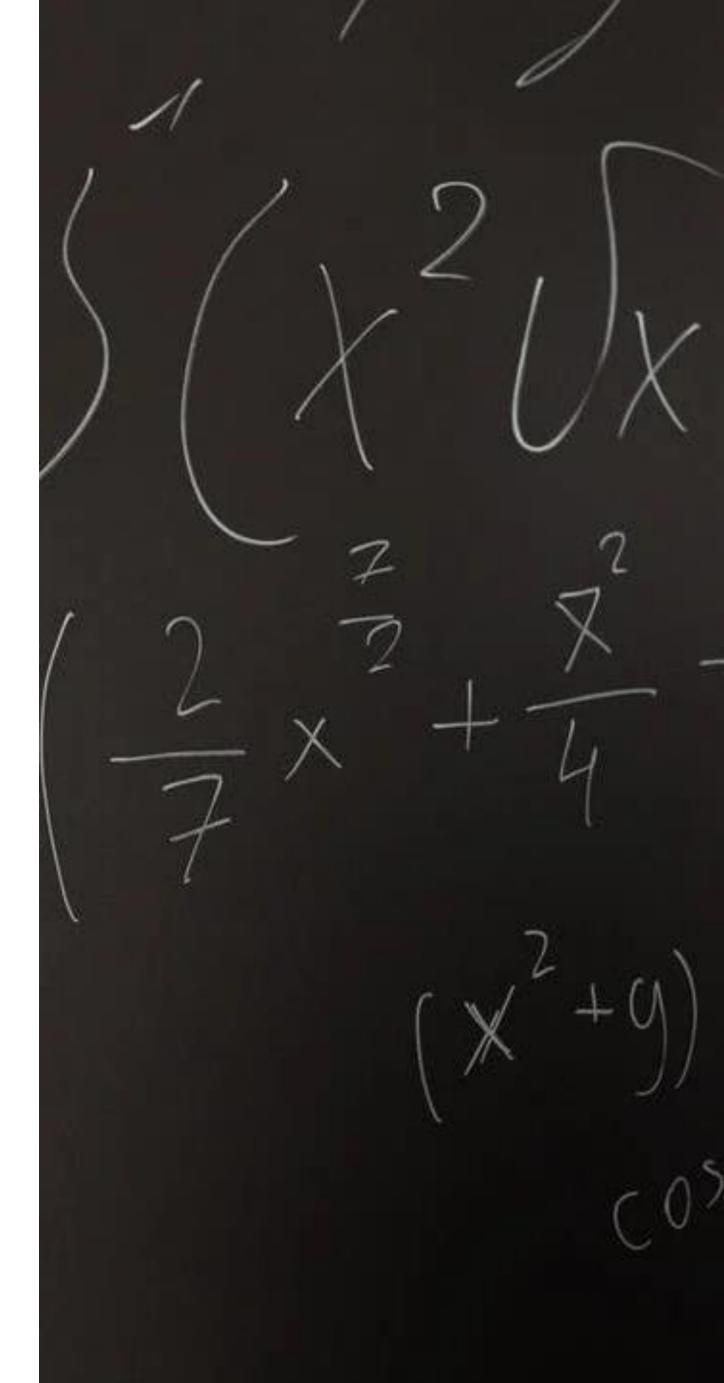
Que és?

- Es un método estadístico que ayuda a definir la estructura de los datos
- Analiza la estructura de las interrelaciones (correlaciones)
 entre un gran número de variables
- Define una serie de dimensiones subyacentes comunes, conocidas como factores
- Permite identificar las dimensiones separadas de la estructura
- Determinar en que grado se justifica cada variable por cada dimensión





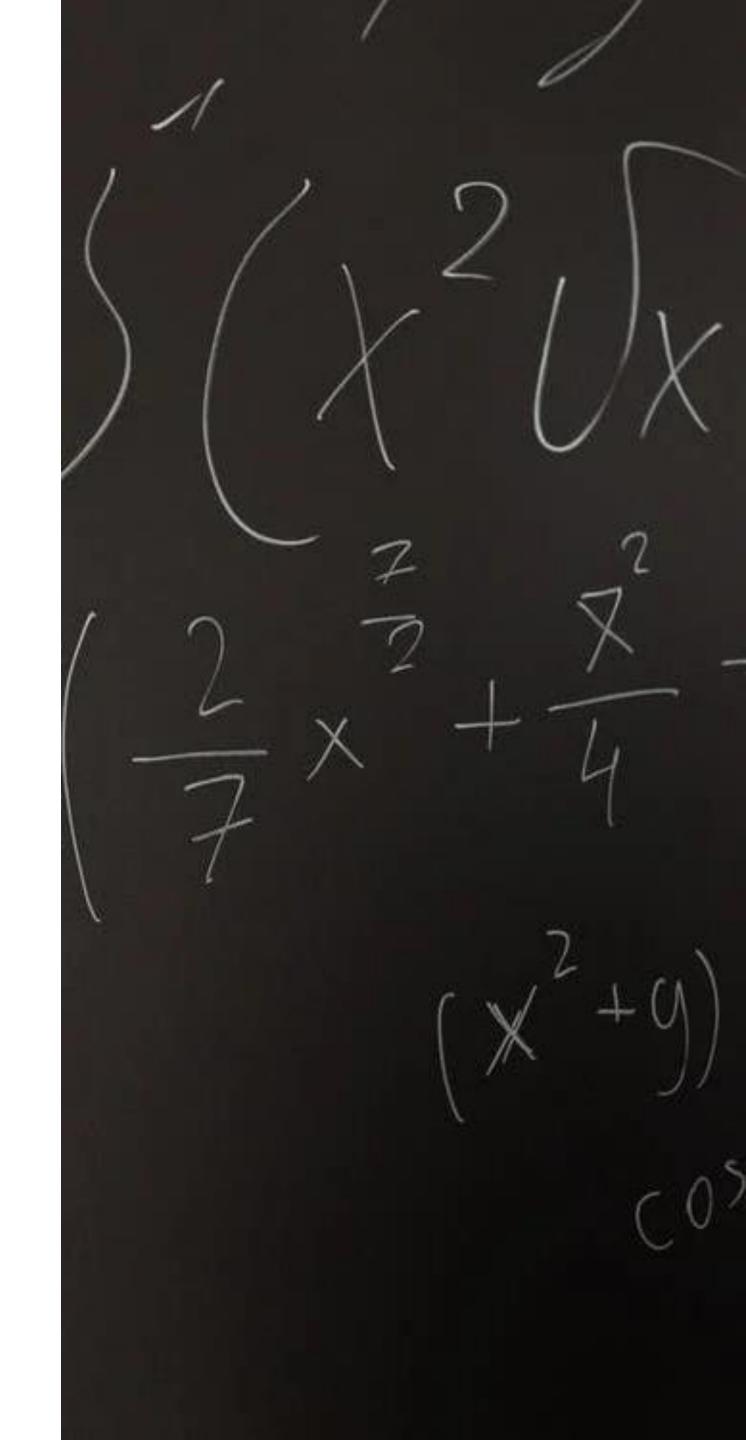






Utilidad

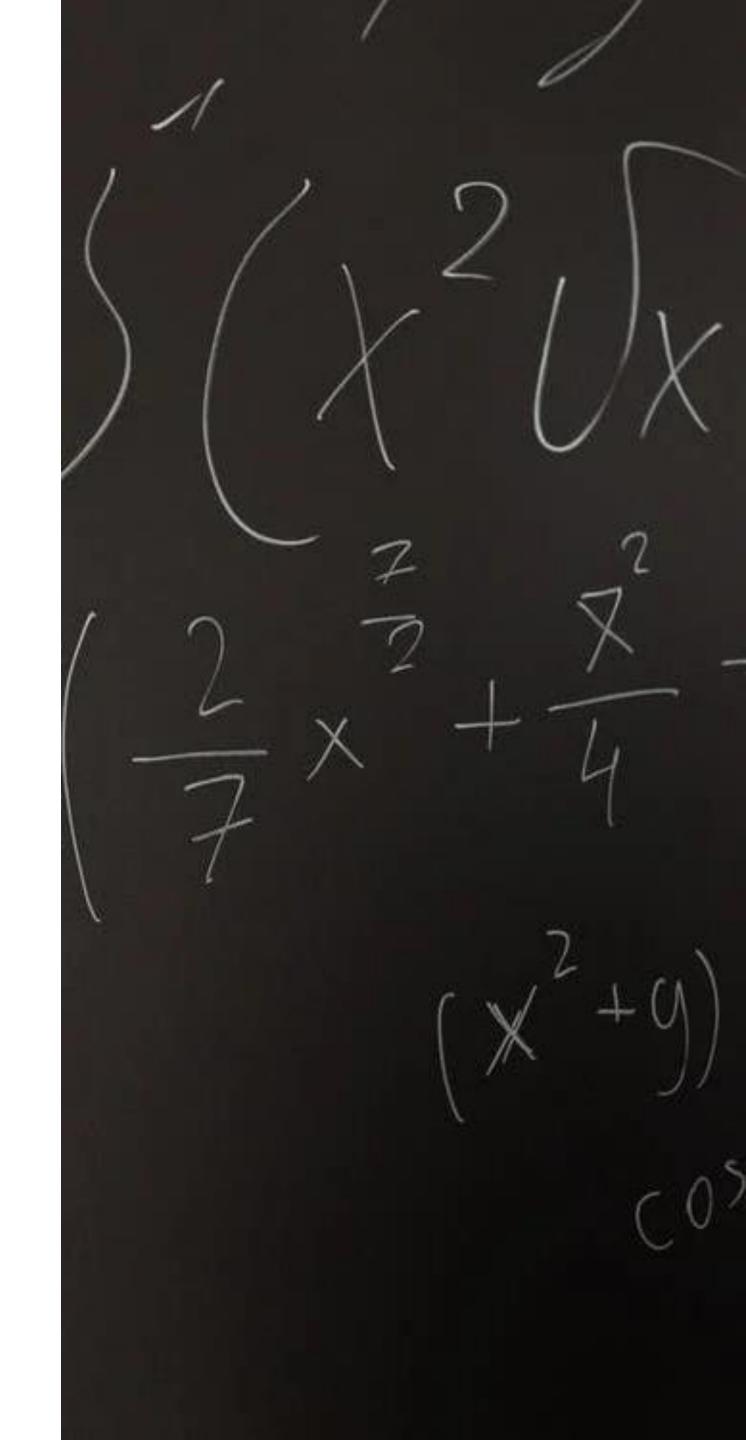
- Resumir y reducir los datos (variables)
- Las dimensiones subyacentes describen los datos con menos variables que las originales
- Se obtiene una puntuación para cada dimensión subyacente y sustituye a las variables originales
- Entre más variables existen aumenta la posibilidad de que las variables no estén correlacionadas y no sean representativas
- Sirve para saber como se relacionan las variables e interpretar mejor los resultados





Diferencia con otros análisis

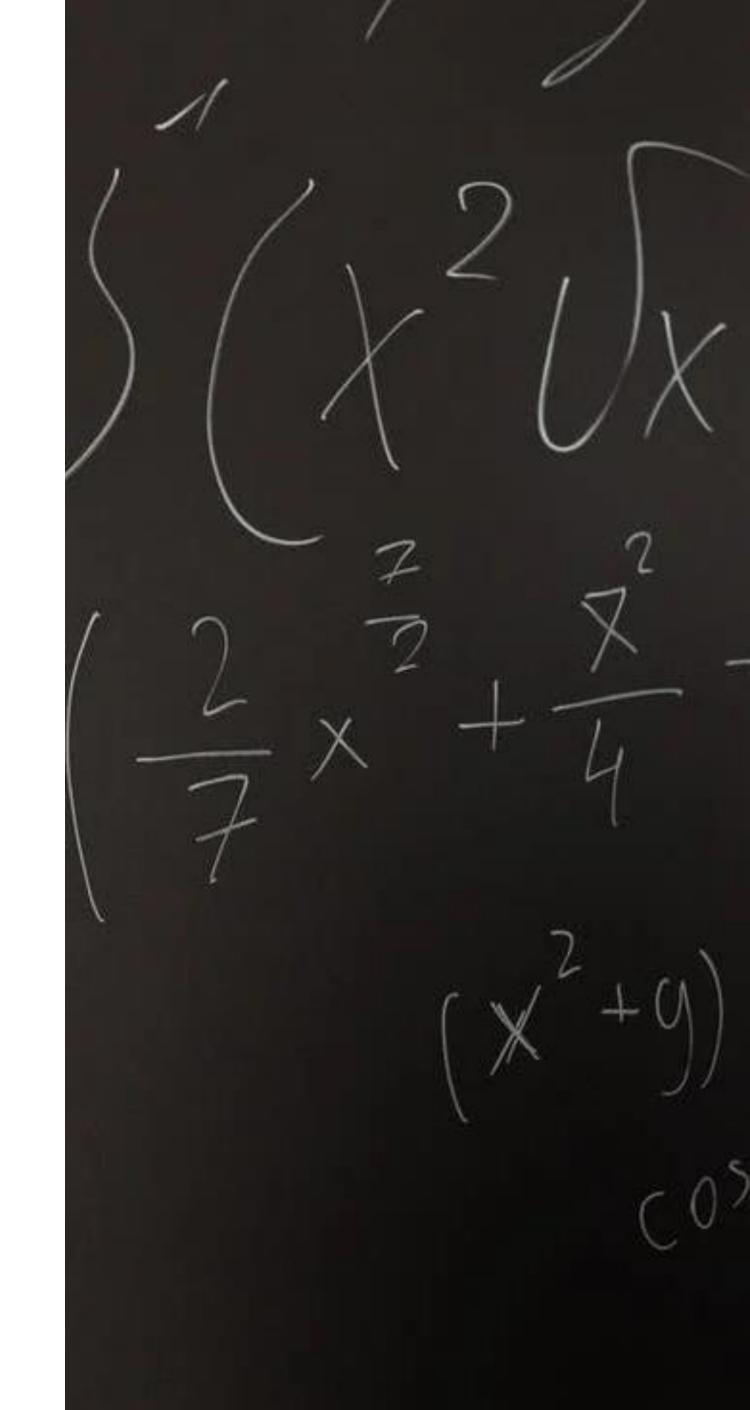
- A diferencia del LDA Análisis Discriminante Lineal y otras técnicas en donde se considera una variable dependiente de otras variables independientes.
- El análisis factorial es una técnica de interdependencia y se consideran todas las variables
- Los valores teóricos (factores) se forman para maximizar su explicación
- No se usa para predecir una variable dependiente
- Interdependencia (identificación de la estructura)





Usos

- Técnica exploratoria
 - "Extrae lo que proporcionan los datos"
 - No hay restricción de la extracción de componentes
- Técnica confirmatoria
 - Qué variables deberían ser agrupadas en un factor
 - Valorar hasta que punto los datos se ajustan a la estructura esperada





Ejemplo

Matriz de correlación

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1 nivel de precios	1								
V2 personal del negocio	0.427	1							
V3 política de retorno	0.302	0.771	1						
V4 disponibilidad del producto	0.470	0.497	0.427	1					
V5 calidad del producto	0.765	0.406	0.307	0.472	1				
V6 profundidad de surtido	0.281	0.445	0.423	0.713	0.325	1			
V7 anchura de surtido	0.354	0.490	0.471	0.719	0.378	0.724	1		
V8 servicio en el negocio	0.242	0.719	0.733	0.428	0.240	0.311	0.425	1	
V9 ambiente del negocio	0.372	0.372	0.774	0.479	0.326	0.429	0.466	0.710	1

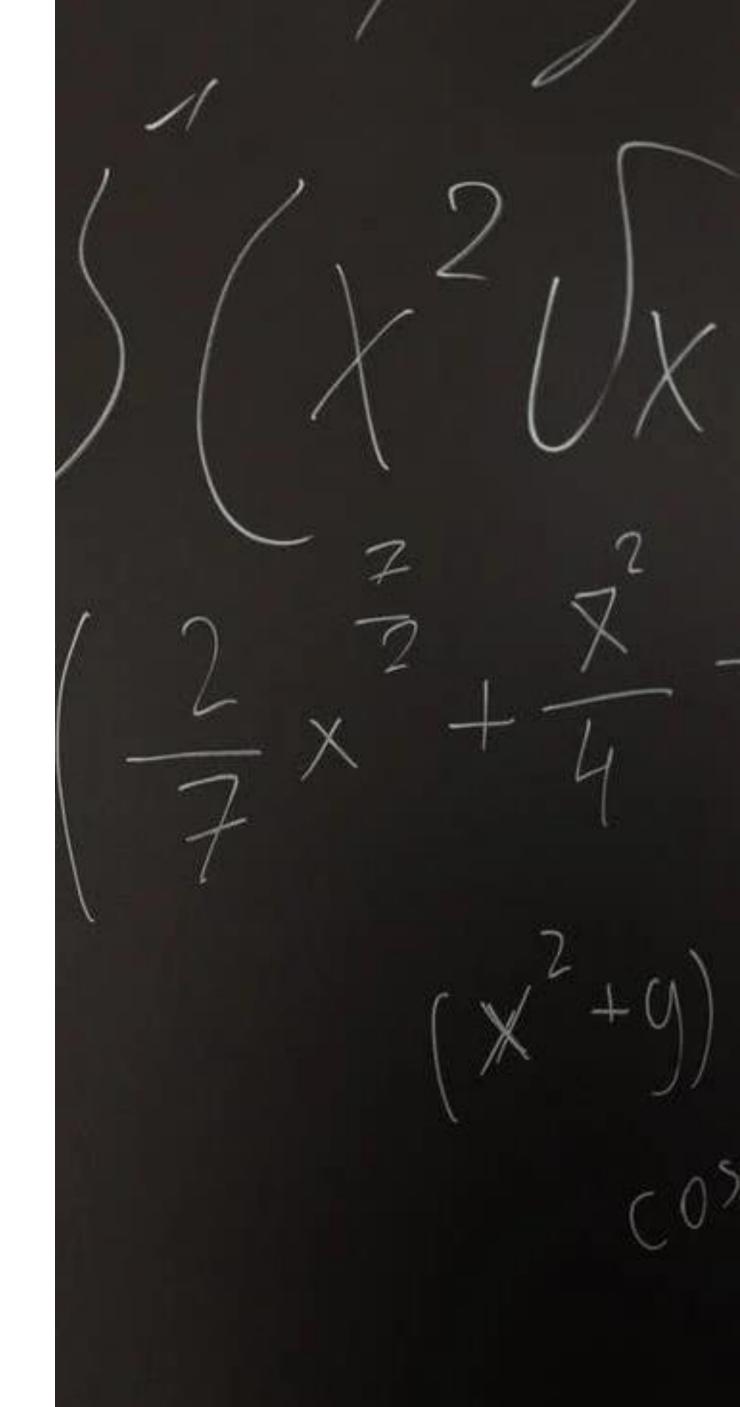
- 4 variables experiencia
- 3 variables variedad producto
- 2 variables precio y calidad

								1	
	V3	V8	V9	V2	V6	V7	V4	V1	V5
V3 política de retorno	1								
V8 servicio en el negocio	0.733	1							
V9 ambiente del negocio	0.774	0.710	1						
V2 personal del negocio	0.741	0.719	0.787	1					
V6 profundidad de surtido	0.423	0.311	0.429	0.445	1				
V7 anchura de surtido	0.471	0.435	0.468	0.490	0.724	1			
V4 disponibilidad del producto	0.427	0.428	0.479	0.497	0.713	0.719	1		
V1 nivel de precios	0.302	0.242	0.372	0.427	0.281	0.354	0.470	1	
V5 calidad del producto	0.307	0.240	0.326	0.406	0.325	0.378	0.472	0.765	1



Proceso

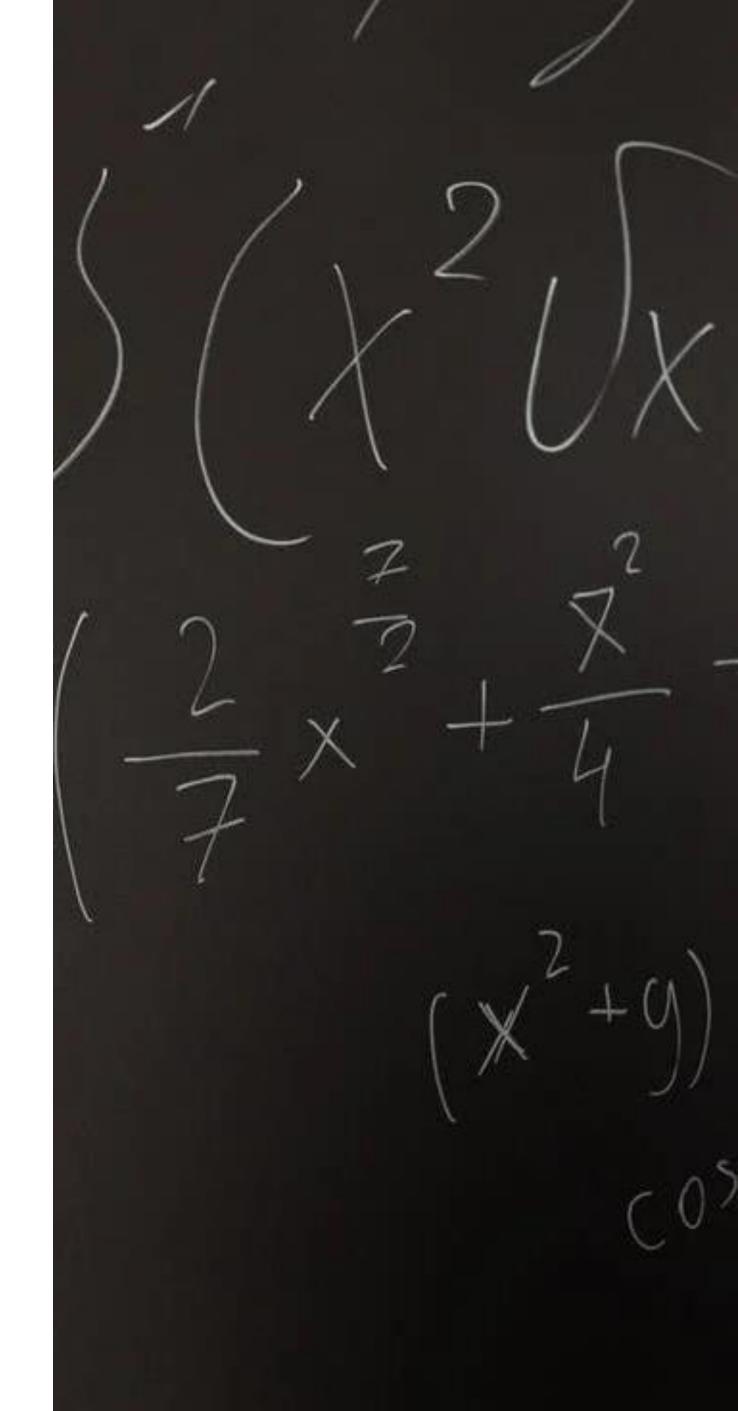
- 1. Objetivo del análisis factorial
- 2. Diseño del análisis factorial
- 3. Supuestos en el análisis factorial
- 4. Estimación de los factores
- 5. Interpretación de los factores
- 6. Validación del análisis factorial





Proceso

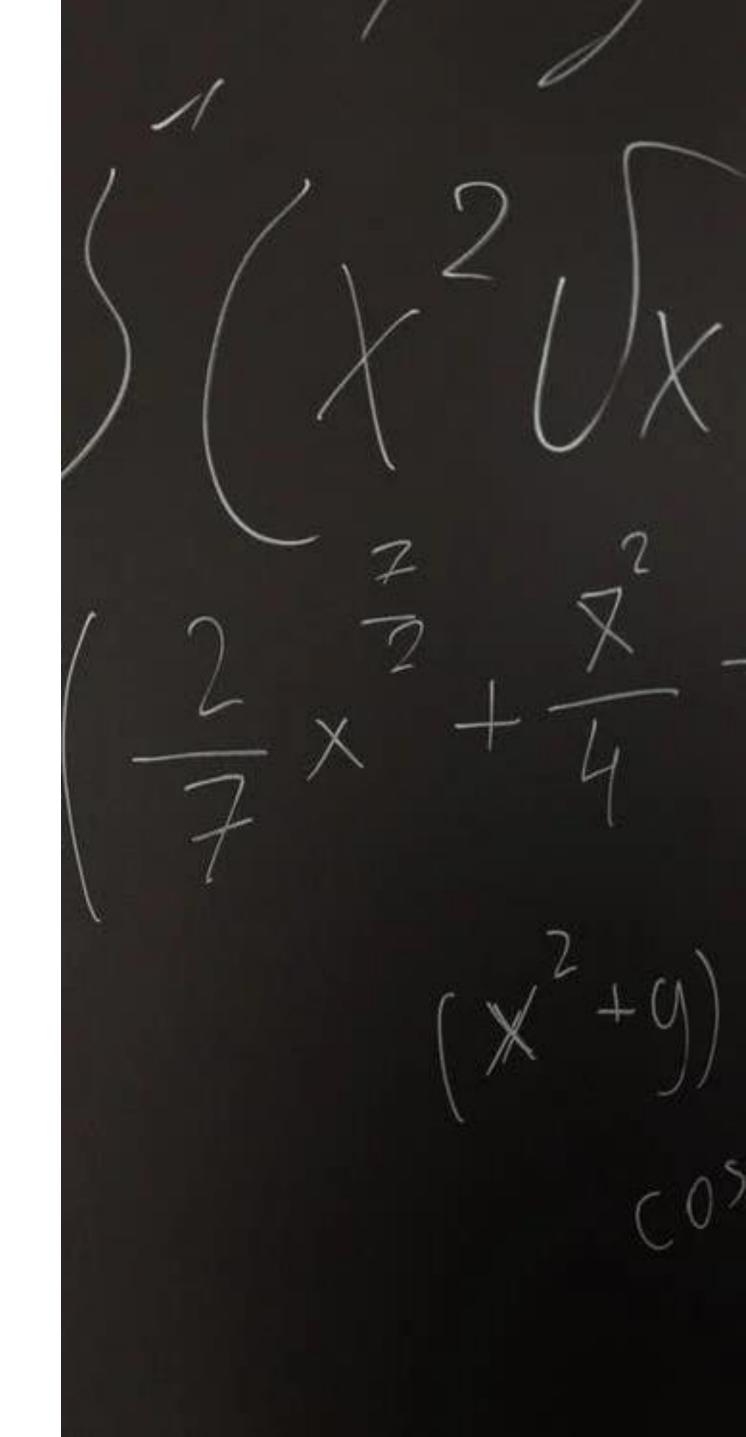
- 1. Objetivo del análisis factorial
- 1.1 Identificación de la estructura resumen de datos
- 1.2 Reducción de datos





1.1 Identificación de la estructura

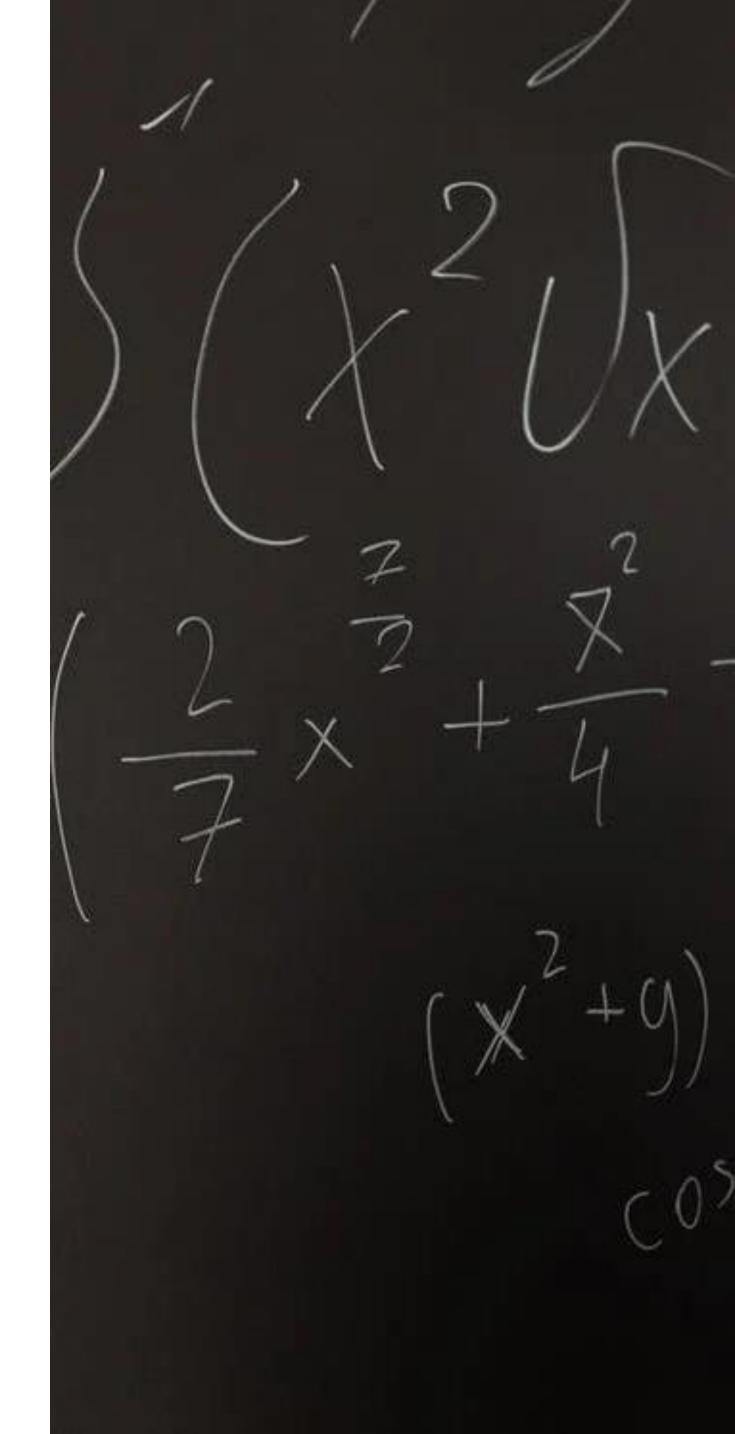
- Identifica correlaciones
- Análisis factorial R
 - Identifica las dimensiones latentes (no fácilmente observables)
- Análisis factorial Q
 - Correlación de los individuos con base a sus características
 - No es muy común (se prefiere conglomerados)





1.2 Reducción de datos

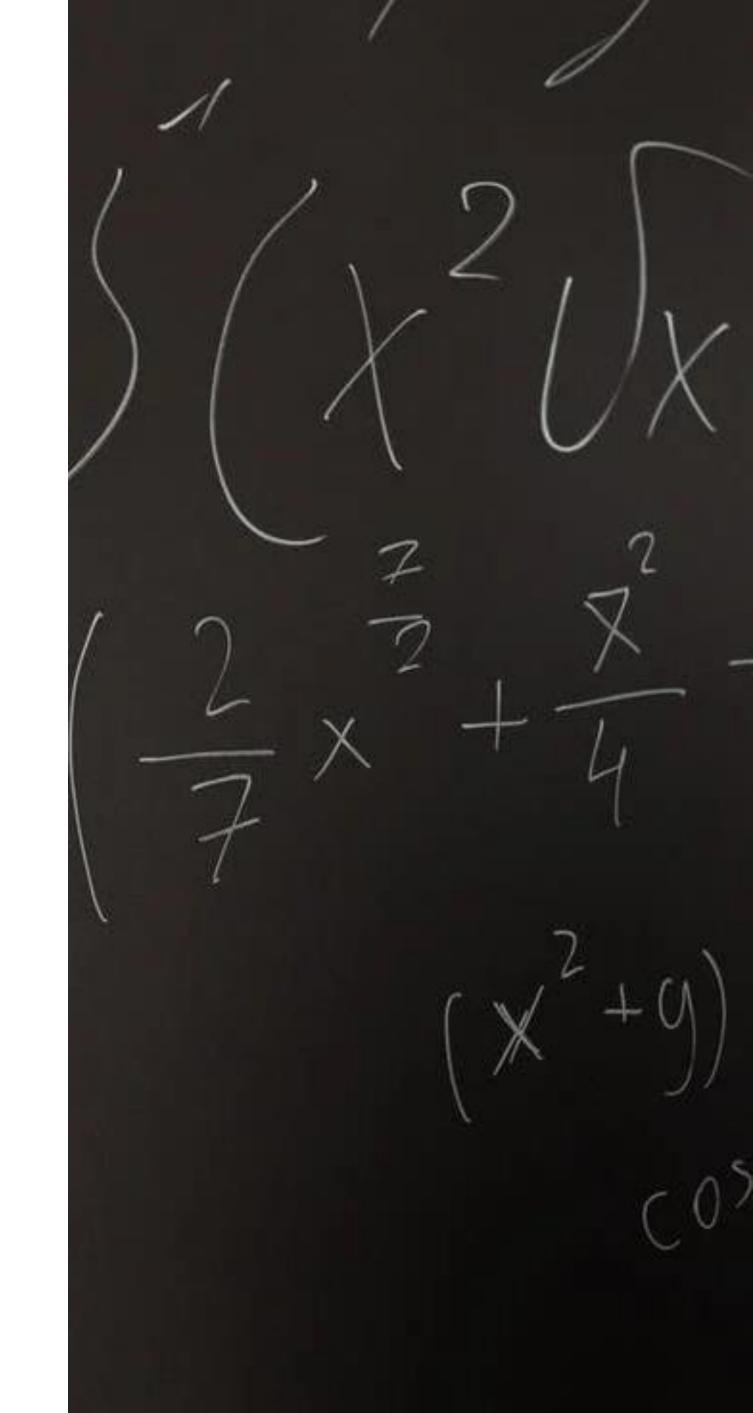
- Crear variables nuevas, más pequeñas.
- Reemplazar parcial o completamente la serie original
- Simplificar el análisis multivariante





Resumen

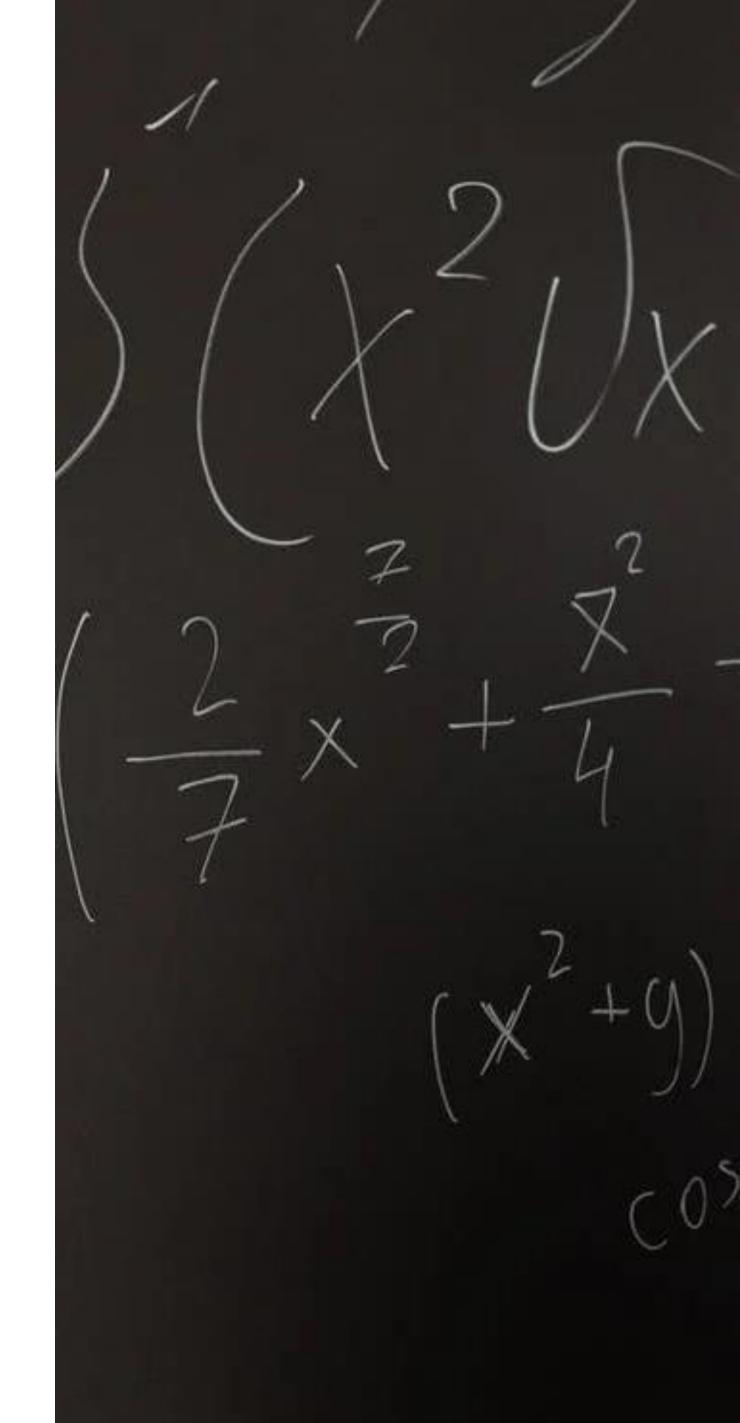
- Muestra interrelaciones entre variables
- Resumen
 - Variables que pueden actuar juntas
 - Variables con mayor impacto
- Reducción de datos
 - Crear nuevas variables que representen los datos
 - Variables reducidas





Selección de variables

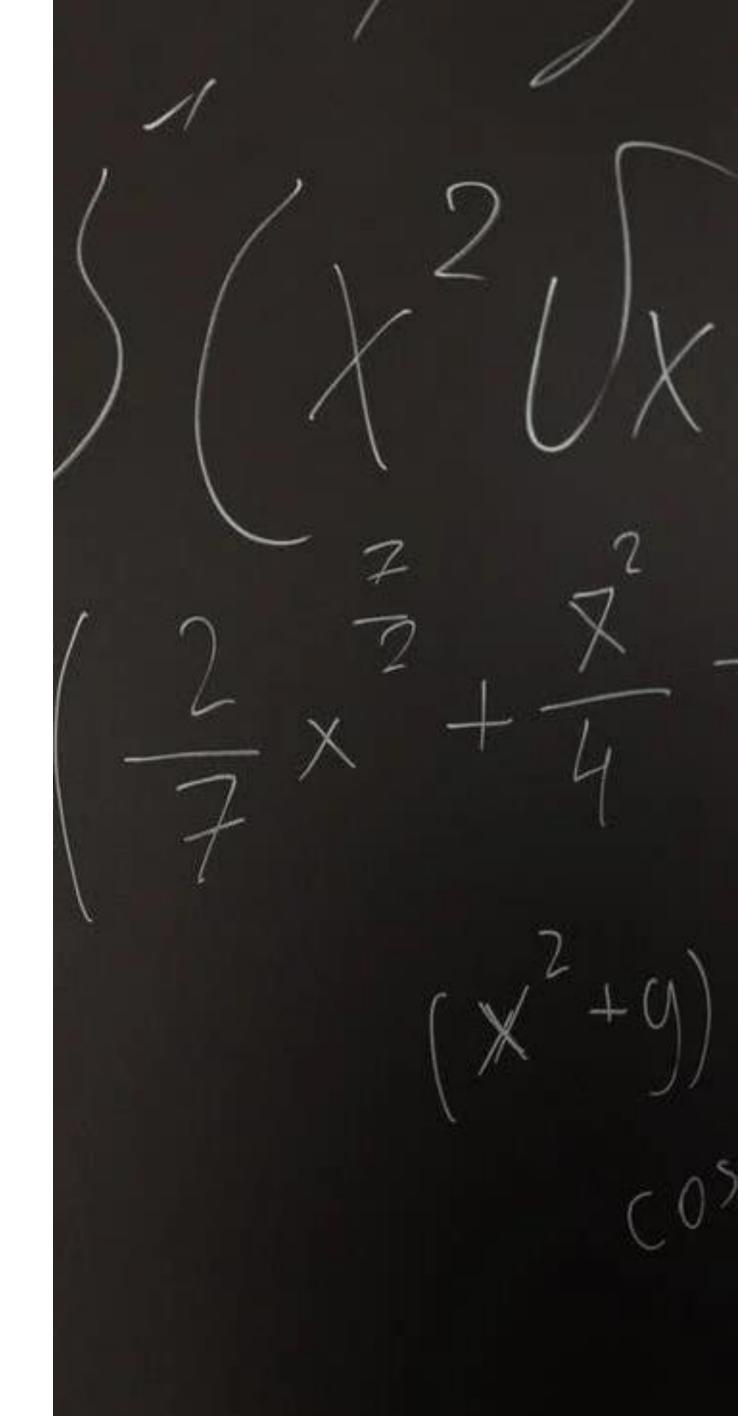
- Trash in , trash out
- Muchas variables pueden producir malos resultados
- Se requiere una base conceptual de cada variable (Negocio)





Proceso

- 2. Diseño del análisis factorial
- 1) El tamaño muestral
- 2) ratio mínima de casos por variable
- 3) la ratio de variables por factor.





Proceso

• 3. Supuestos del análisis factorial

Normalidad

Homocedasticidad (la varianza del error condicional a las variables explicativas es constante a lo largo de las observaciones.)

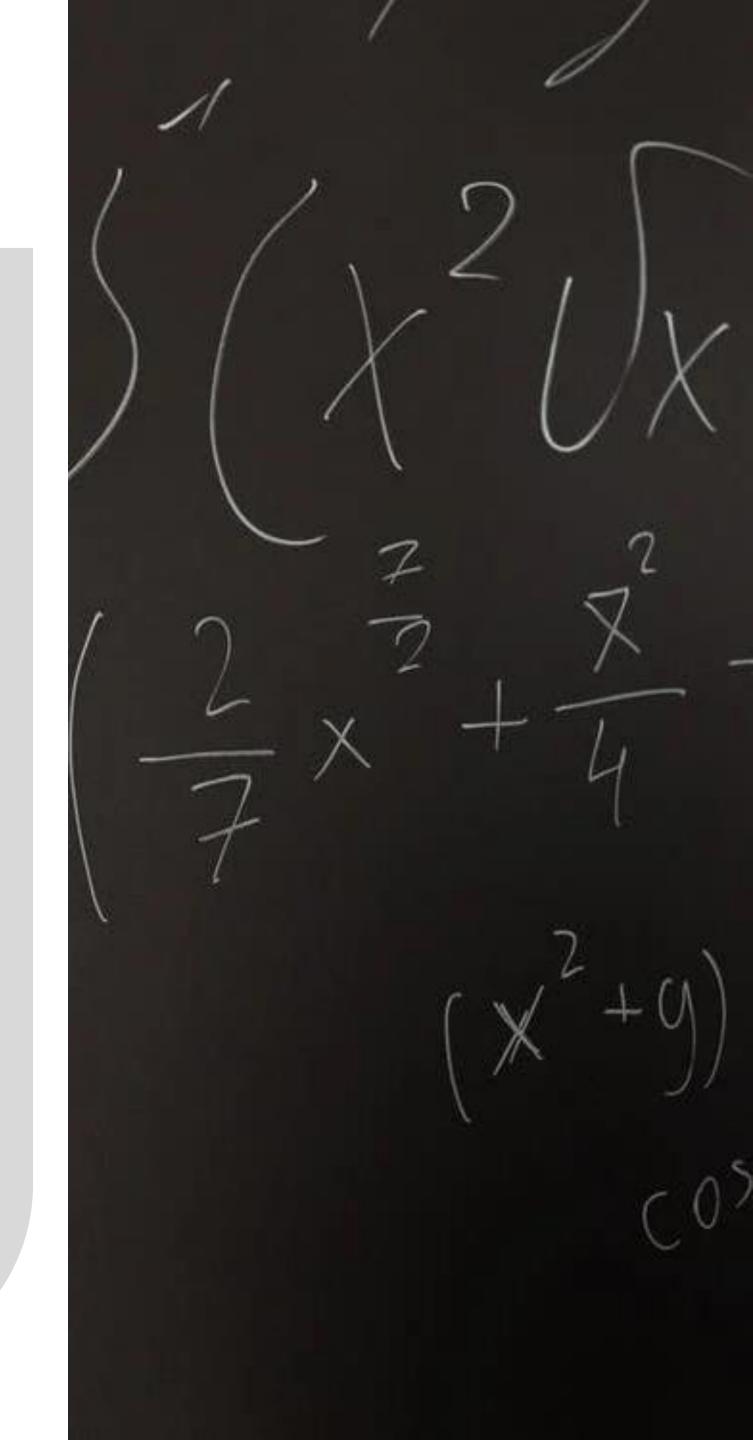
Linealidad (dependencia lineal)

Matriz de correlaciones estadisticamente significativas (no hay azar)

Contraste de Bartlett - Probar si k muestras provienen de poblaciones

con la misma varianza

Medida de Adecuación Muestral (MSA) - contrasta si las correlaciones parciales entre las variables son pequeñas



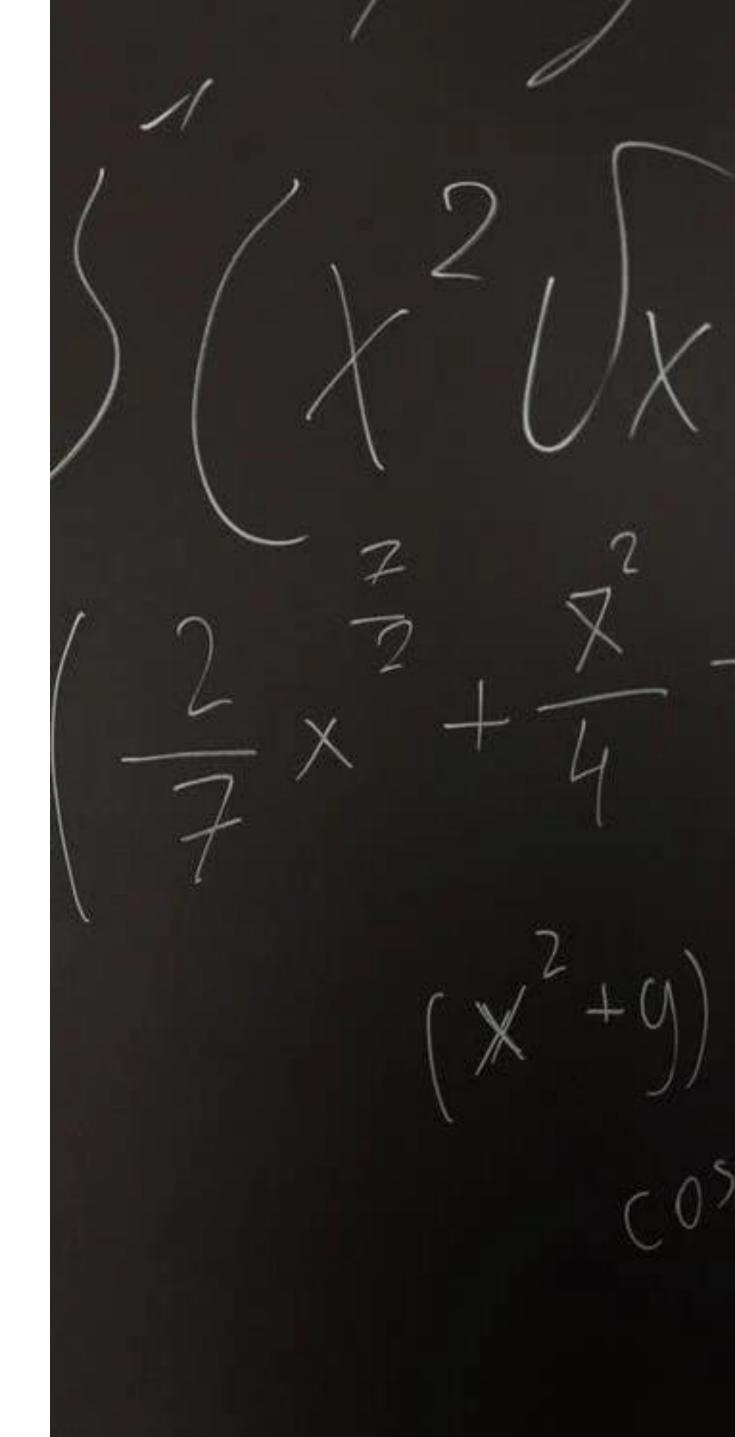


Proceso

4. Estimación de factores

Matriz de correlación a matriz de factores

Seleccionar el número de componentes que se van a mantener

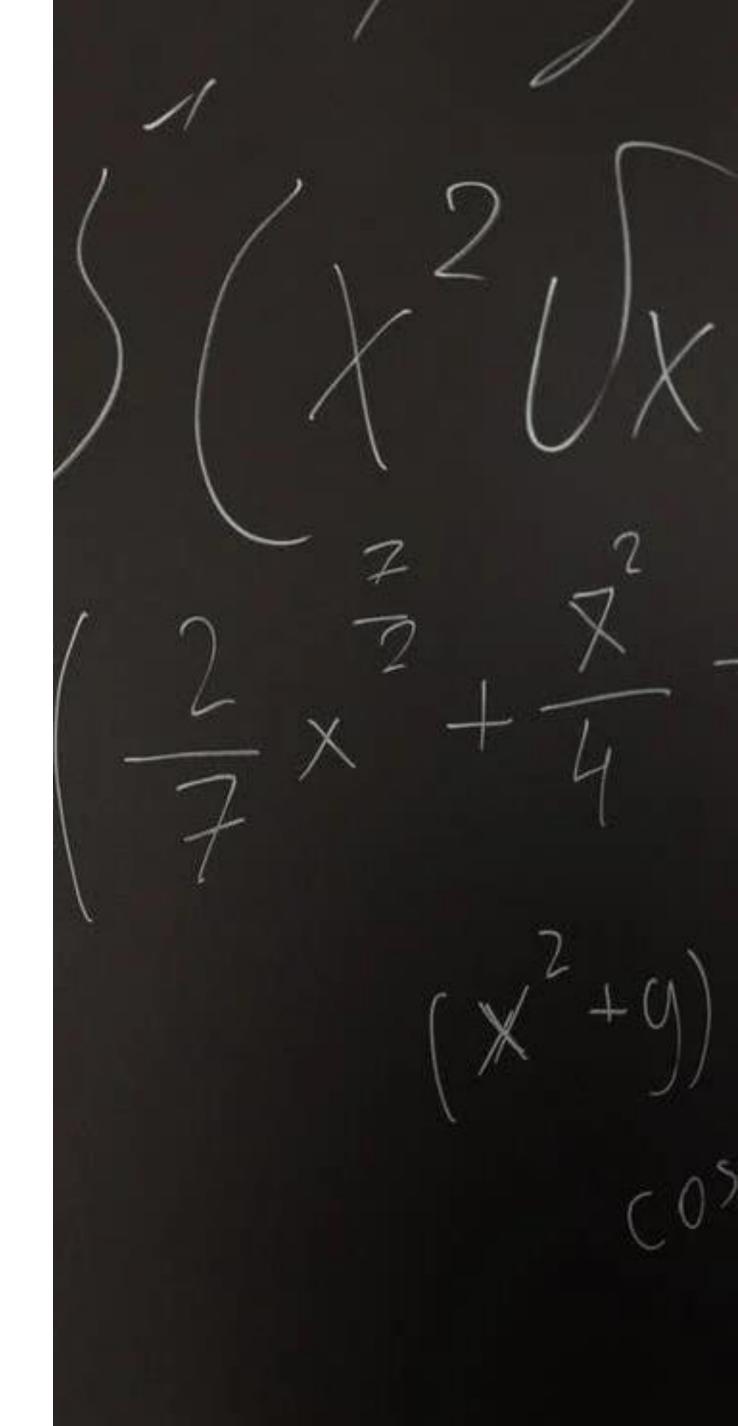




Proceso

• 5. Interpretación de los factores

A nivel negocio se explica



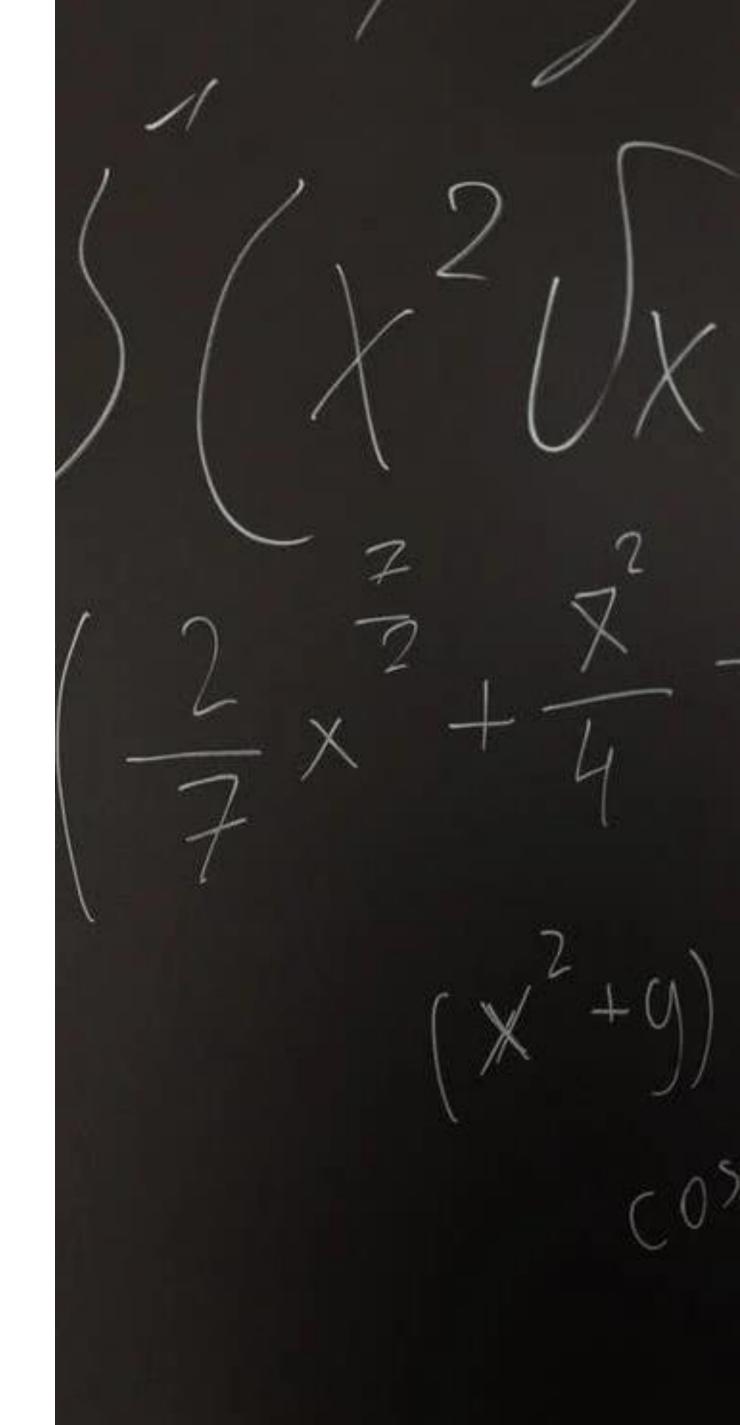


Proceso

• 6. Validación del análisis factorial

Analisis de división de la muestra

Aplicación de muestras nuevas



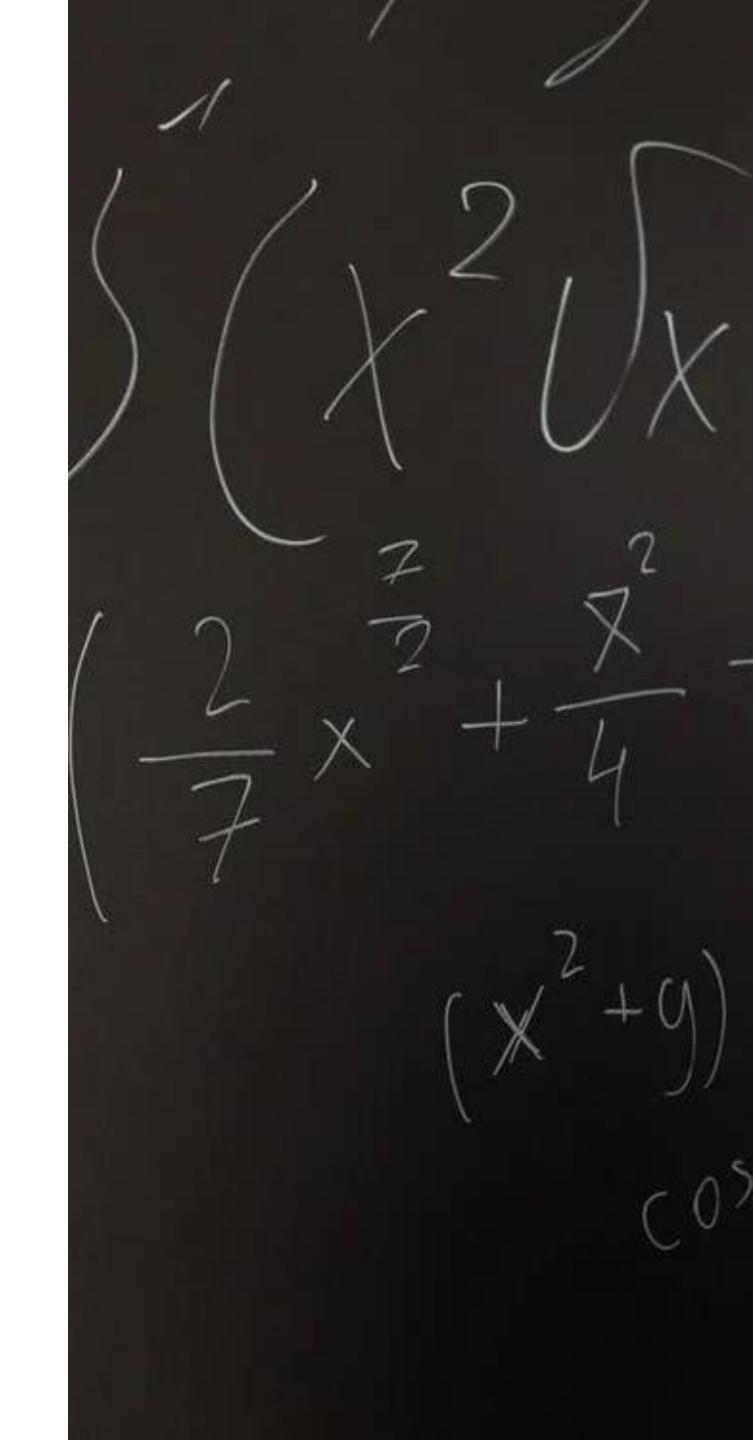


Modelo

- Método de factores:
 - Varianza total
 - Varianza comun

Valor diagonal	Va	arianza
Unidad	Varia	anza total
Comunalidad	Común	Específica y error
	Varianza extraída Varianza perdida	

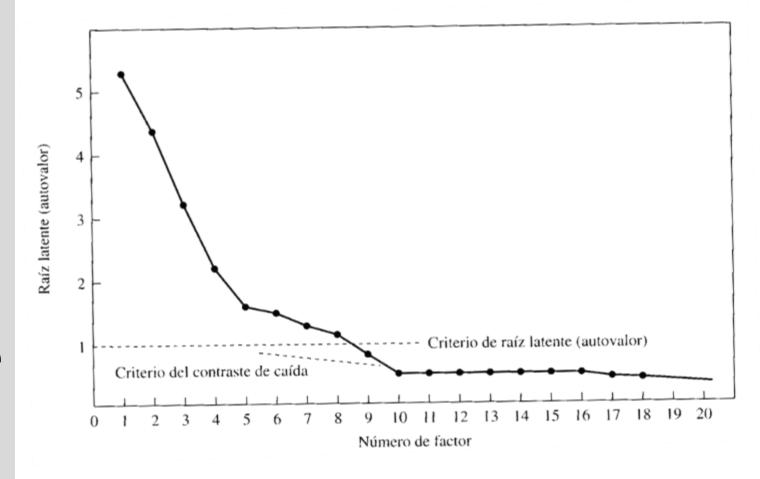
- Número de factores a ser retenidos
 - Raíz latente cualquier factor debe explicar al menos una variable
 - Criterio a priori el investigador sabe cuantos factores
 - Porcentaje de la varianza % de la varianza total extraída
 - Criterio de contraste de caída antes de que la varianza empiece a dominar la estructura

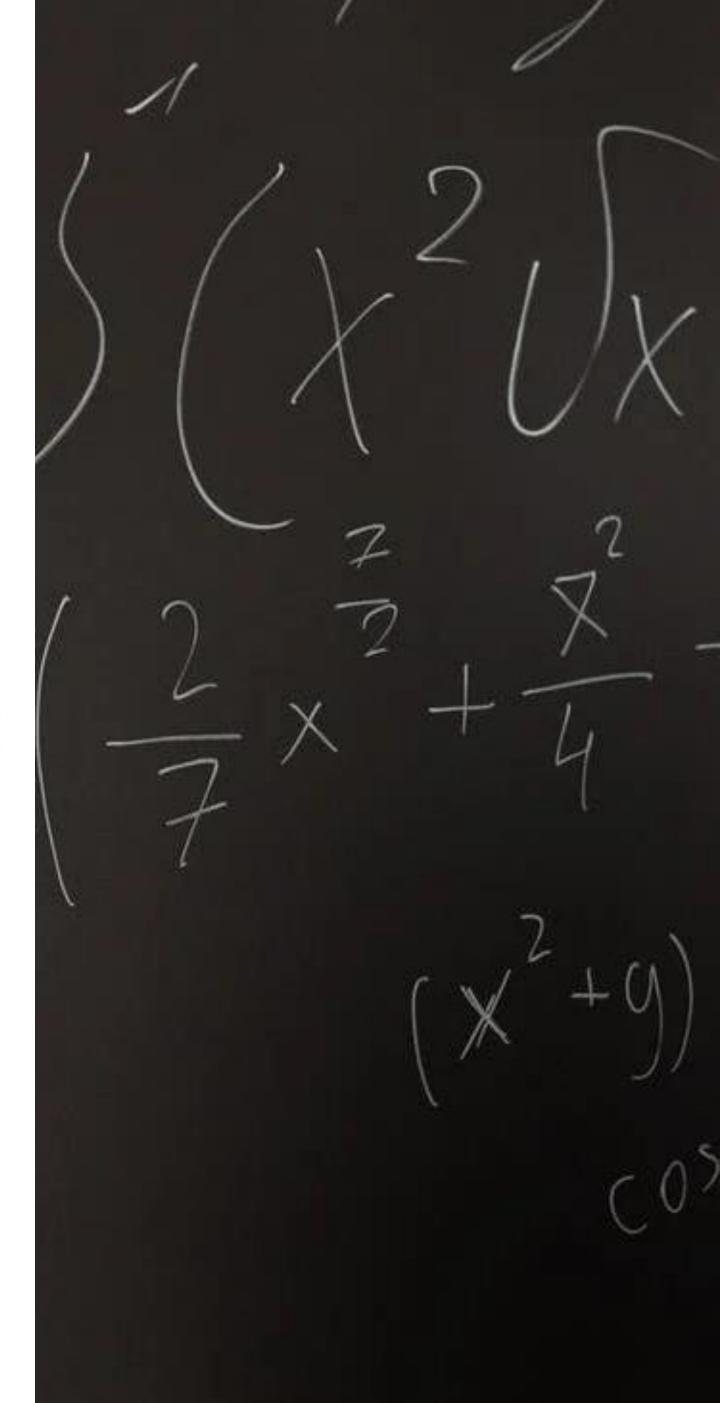




Modelo

- Número de factores a ser retenidos
 - Raíz latente cualquier factor debe explicar al menos una variable
 - Criterio a priori el investigador sabe cuantos factores
 - Porcentaje de la varianza % de la varianza total extraída
 - Criterio de contraste de caída antes de que la varianza empiece a dominar la estructura

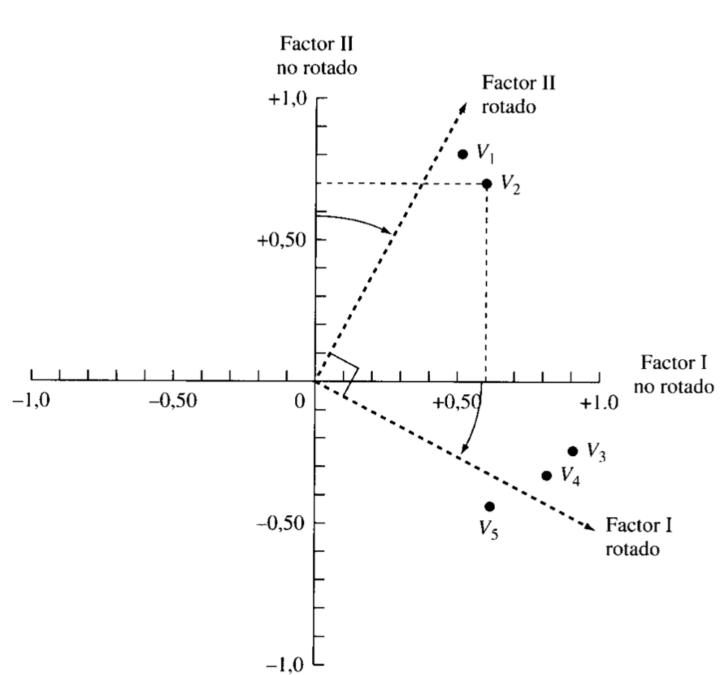


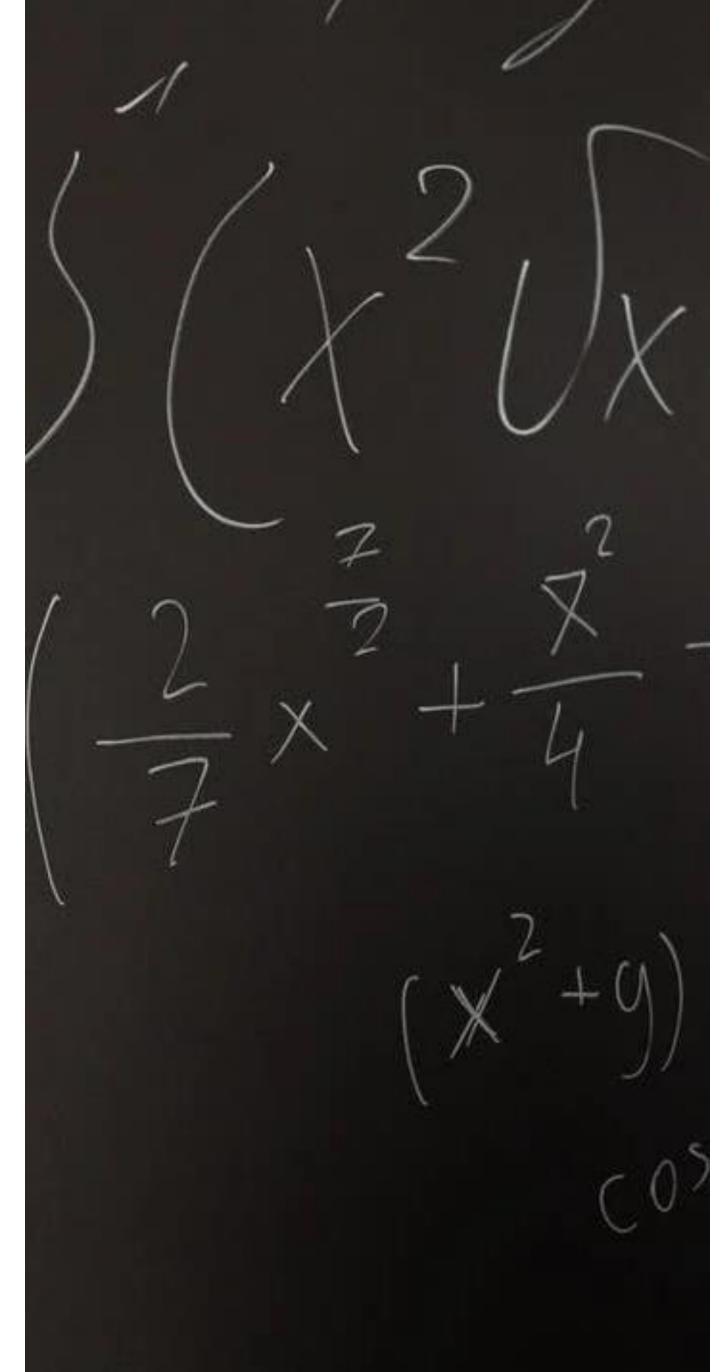




Modelo

 Rotación de factores: se giran los origenes de los ejes de referencia de los factores, el rotar la matriz de factores es redistribuir la varianza para lograr un patrón de factores más simple.







- Dentro de Colab
 - https://colab.research.google.com

```
require File.expand_pott
# Prevent database trace
 abort("The Rails environment to
 require 'spec_helper'
  require 'rspec/rails'
  require 'capybara/rssec
   require 'capyboro/reils'
   Capybara.javascript
   Category.delete_all; Category.
    Shoulda::Matchers.com
      config.integrate (a)
        with.test_fromework
        with.library :roils
      # Add additional res
       # Requires supporting
        # spec/support/ and the
        # run as spec files by
         # run twice. It is
```

Recursos



Bibliográficos

- Hair, Anderson, Tatham, Black, Análisis Multivariante, Prentice Hall Iberia, 1999
- Irini Mavrou, Análisis factorial exploratorio, Universidad Antonio de Nebrija, ISSN 1699-6569, https://www.nebrija.com/revista-linguistica/analisis-factorial-exploratorio.html